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

Doc Type: **Working Group Document**
Title: **Unicode Technical Report #3: Early Aramaic, Balti, Kirat (Limbu), Manipuri (Meitei), and Tai Lü scripts**
Source: **Rick McGowan (UTC), Michael Everson (IE)**
Status: **Discussion document**
Action: **For consideration by JTC1/SC2/WG2**
Date: **1999-07-20**

Apparently, Unicode Technical Report #3, published in 1993, was never submitted to JTC1/SC2/WG2 for distribution. Since a number of scripts in that Technical Report have already been encoded, or since other proposals for them have already been put forward to WG2, distribution of the UTR#3 itself at the present time would not be very useful.

This document presents the names and code tables of the scripts in the UTR#3 which have not yet been otherwise discussed in WG2, namely: Early Aramaic, Balti, Kirat/Limbu, Manipuri/Meitei, and Tai Lü.

These proposals were written by Rick McGowan. Formatting and some minor editing in this document is by Michael Everson.

Early Aramaic

The Aramaic alphabet branched from the 22 letter alphabet used for Phoenician and evolved along separate lines culminating in Syriac, Arabic and other scripts. The Early Aramaic block should be used for Late Aramaic (especially papyri), Palmyrene, and Nabataean, Mandaic and their immediate precursors and successors.

The order shown in the accompanying chart matches the order of the Early Phoenician block and the shapes shown there are in the Palmyrene style.

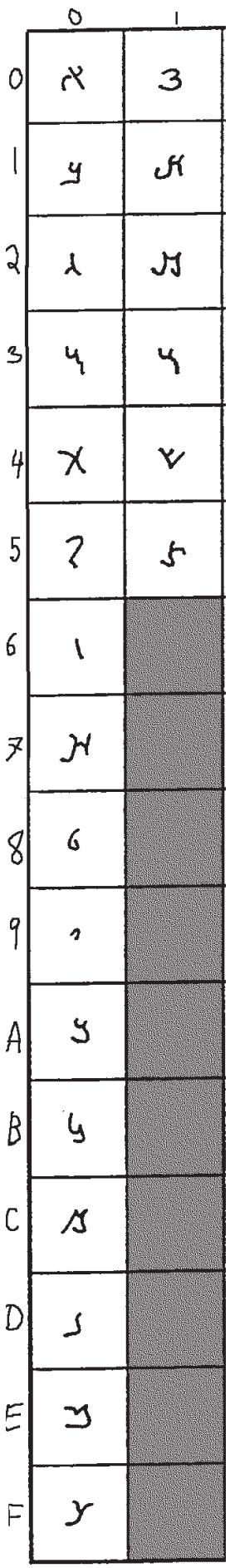
See the Phoenician block introduction and the Early Alphabets block introduction in UTR#3 for further information and issues.

Healey, John F. *The Early Alphabet.*
Cross, Frank Moore. *The Invention and Development of the Alphabet.*
Diringer, David. *Writing.*

Aramaic Names List, draft 1999-07-20

```
00 ARAMAIC LETTER ALEPH
01 ARAMAIC LETTER BETH
02 ARAMAIC LETTER GIMEL
03 ARAMAIC LETTER DALETH
04 ARAMAIC LETTER HE
05 ARAMAIC LETTER WAW
06 ARAMAIC LETTER ZAIN
07 ARAMAIC LETTER HETH
08 ARAMAIC LETTER THET
09 ARAMAIC LETTER YODH
0A ARAMAIC LETTER KAPH
0B ARAMAIC LETTER LAMED
0C ARAMAIC LETTER MEM
0D ARAMAIC LETTER NUN
0E ARAMAIC LETTER SAMEKH
0F ARAMAIC LETTER AIN

10 ARAMAIC LETTER PE
11 ARAMAIC LETTER SAN
12 ARAMAIC LETTER QOPPA
13 ARAMAIC LETTER RESH
14 ARAMAIC LETTER SHIN
15 ARAMAIC LETTER TAU
```



Balti

The Balti script is now extinct, but was formerly used to write the Balti language of Baltistan, in what is now part of Ladakh in Northern Kashmir. The script was apparently introduced in about the fifteenth century CE when the people converted to Islam. It is related to the Arabic script.

In contrast to many other Brahmic scripts, Balti is written from *right to left* horizontally, in the Arabic manner. All of the vowel signs except long *a* are integrated into the glyphs used for consonants, becoming projections from the consonants rather than being separate marks as in most of the modern Brahmic scripts. The consonants apparently have an inherent *a* vowel (or an explicit vowel sign *a* may appear; there may not be a distinction between long and short *a*). There appears to be a sign (overdot) used to indicate the end of a word, but no interword spacing seems to be used.

The base form of *b* is the same as *p* and *t*; only the dots distinguish these. There are two other similar pairs. These appear to approximately parallel similar dotted versus dotless letters in Arabic.

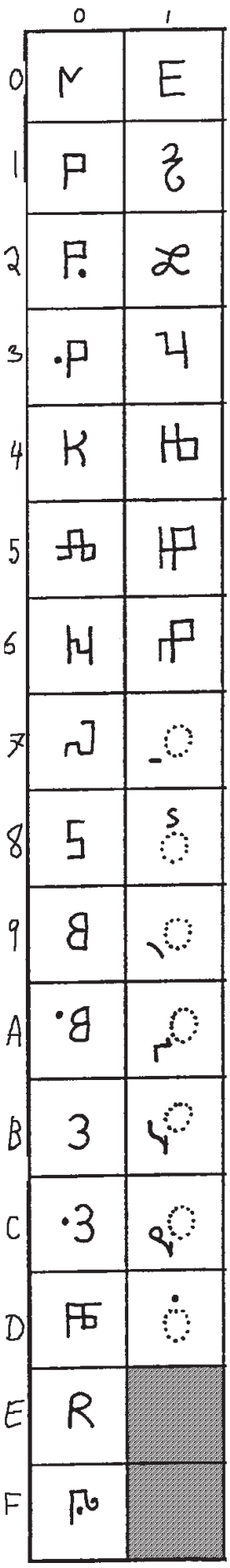
Issues: The set of Balti consonants is too small to make it worth encoding parallel to any of the other Brahmic scripts, or to Arabic. Not enough information is available at this time to determine the completeness of the accompanying chart. The digits, if any, are unknown. It is unknown how much literature is available in the old Balti script, or what the level of scholarly interest in it is. The function of the character listed in the names list as “Balti null vowel or word ending” is uncertain.

Grierson, G. *A. Linguistic Survey of India, Vol. 3.*
One photocopy of 2 pages (326 and 327) from an unknown volume in German.

Balti Names, draft 1999-07-20

```
00 BALTI LETTER A
01 BALTI LETTER BA
02 BALTI LETTER PA
03 BALTI LETTER TA
04 BALTI LETTER GA
05 BALTI LETTER HHA
06 BALTI LETTER CA
07 BALTI LETTER CHA
08 BALTI LETTER DA
09 BALTI LETTER RA
0A BALTI LETTER ZA
0B BALTI LETTER SA
0C BALTI LETTER SHA
0D BALTI LETTER KA
0E BALTI LETTER LA
0F BALTI LETTER MA
```

```
10 BALTI LETTER NA
11 BALTI LETTER HA
12 BALTI LETTER JA
13 BALTI LETTER KHA
14 BALTI LETTER THA
15 BALTI LETTER TSA
16 BALTI LETTER NGA
17 BALTI VOWEL SIGN A
18 BALTI VOWEL SIGN AA
19 BALTI VOWEL SIGN E
1A BALTI VOWEL SIGN I
1B BALTI VOWEL SIGN O
1C BALTI VOWEL SIGN U
1D BALTI NULL VOWEL OR WORD ENDING?
```



Kirat (Limbu)

The Limbu (or Kirat or Kiranti) alphabet is (or was) used among the Limbu of Sikkim and Darjeeling. Kirat is structurally similar to the Lepcha (Rong) script. It has 20 consonants (including the stand-alone "A" as in other Brahmic scripts), 8 vowel signs, 7 (or 8 or 10?) final consonants. Letters YA, RA, and WA may be subscripted in a manner similar to the Tibetan and Rong scripts. There appears to have been, at sometime in the past, an orthographic reform, and two slightly different varieties of the script appear to be in existence.

There are three other symbols needed for proper pronunciation of Limbu. These are *mukphreng* (aspiration mark), *kehmphreng* (length mark) and *sa-i* (possibly the virama). The *sa-i* appears to be used to remove the inherent *A* sound like a *virama. Sa-i* has been conjectured to occur visibly only in word-medial position. It has been observed also in apparent word-final position. Its function may be therefore different from an invisible *virama.*

Kirat appears to include three other marks, the names of which are not presently known. These are (1) a mark indicating colon or full stop, (2) a mark indicating a prolonged final note during a chant, (3) a mark which looks like the Oriya *anusvara* (a circle above) indicating an acute type of accent.

The accompanying chart was prepared from a draft supplied by Lloyd Anderson. The ISCII model and layout is followed in the accompanying chart. The shaded cells to the far right are final consonants (lower nine cells), a "tr" conjunct and a "j" rendering form.

Issues: It is not known whether the Kirat script is still in use as of this writing (1992). It was reported in 1855 as nearly extinct, but sources as recent as 1979 are available.

This draft for Kirat is by no means complete. Sources vary even as to the correct number of final consonants (or "conjoint letters" called *kedumba sok*); there may be as many as ten of them.

There are two different approaches to encoding of Kirat. If the script is postulated to contain an invisible *virama* distinct from *sa-i*, then the final consonants could be rendered in text by using this *virama* followed by the corresponding normal forms If, however, no such invisible virama is postulated, then the final consonants should be encoded distinctly. There is no concrete evidence yet available [to this author] for or against such an invisible *virama* that is distinct from *sa-i.* Both are transliterated into Devanagari by use of half-consonant forms, as Devanagari has no such distinction at all. The final consonants cannot be rendered alone by use of *sa-i*, since the *sa-i* appears to be always visible when it occurs, and *kedumba sok* forms also occur without the *sa-i.* There thus appears to be some distinction, and *sa-i* alone is insufficient to generate both forms. *Sa-i* is also seen with full consonants, where it presumably functions like a *virama* (in eliding the inherent vowel).

In either case, the script bears some similarity to the Lepcha script, and it seems that the same conceptual model should be used for both. Kirat could be laid out in a manner compatible with ISCII and parallel to Devanagari as far as the arrangement of its vowels and consonants. However, since it has a somewhat smaller complement of consonants than Devanagari, and needs no precomposed long vowels, many empty codepoints are unnecessarily scattered throughout such an encoding. Kirat could also be encoded parallel to Tibetan as far as the arrangement of its consonants.

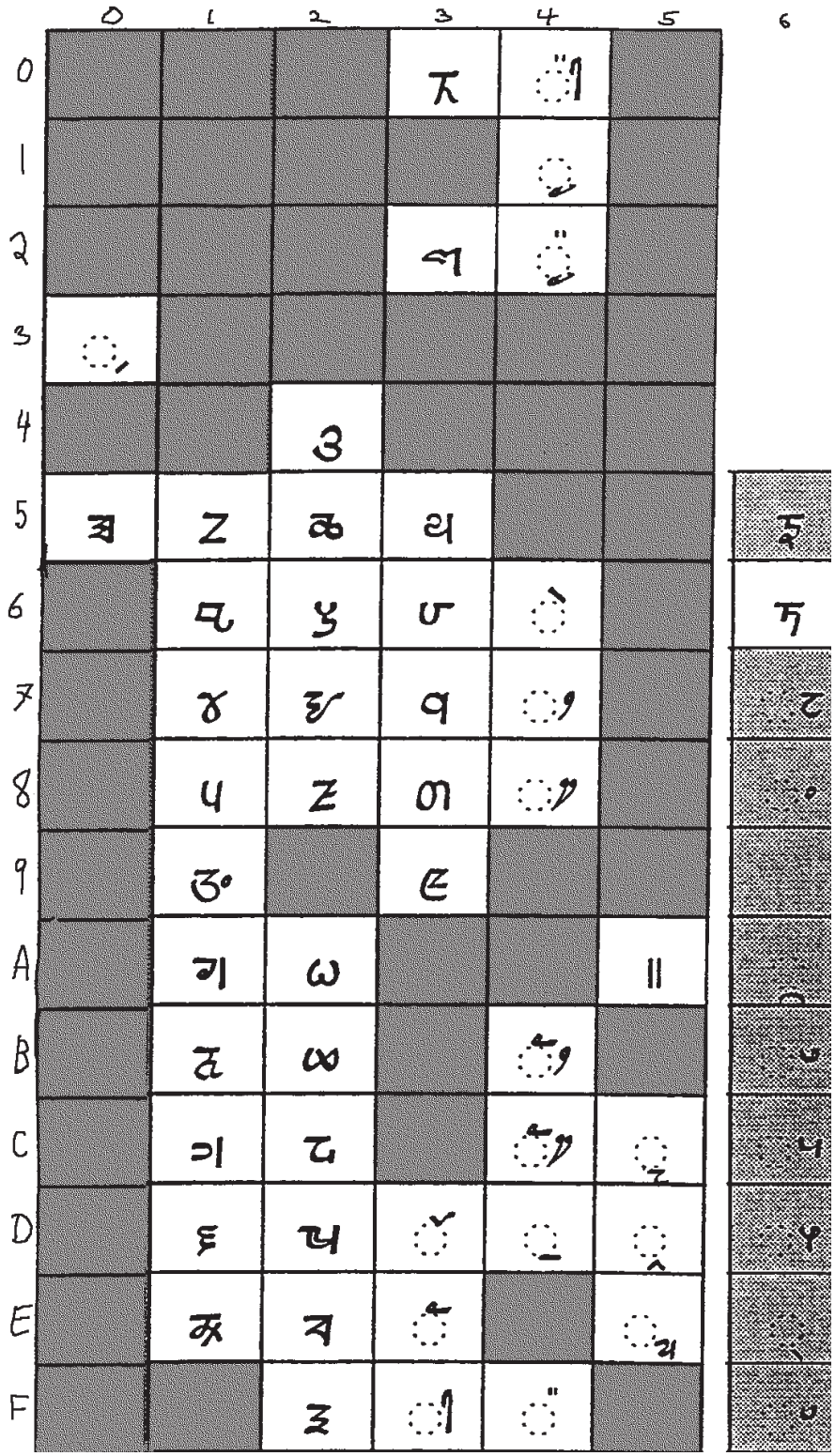
Campbell, *A. Note on the Limboo Alphabet of the Sikkim Himalaya.*
Chemsong, Iman Singh. *The Kirat Grammar (Limbu).*
Subba, B. B. *Limbu Nepali English Dictionary.*
Kirat Primary Book.
Limbu Reader VI.

Kirat (Limbu) Names List, draft 1999-07-20

00 Reserved
01 Reserved
02 Reserved
03 KIRAT ASPIRATION MARK (mukhphreng)
04 Reserved
05 KIRAT LETTER A
06 Reserved
07 Reserved
08 Reserved
09 Reserved
0A Reserved
0B Reserved
0C Reserved
0D Reserved
0E Reserved
0F Reserved
10 Reserved
11 Reserved
12 Reserved
13 Reserved
14 Reserved
15 KIRAT LETTER KA
16 KIRAT LETTER KHA
17 KIRAT LETTER GA
18 KIRAT LETTER GHA
19 KIRAT LETTER NGA
1A KIRAT LETTER CHA
1B KIRAT LETTER CHHA
1C KIRAT LETTER JA
1D KIRAT LETTER JHA
1E KIRAT LETTER NYA
1F Reserved
20 Reserved
21 Reserved
22 Reserved
23 Reserved
24 KIRAT LETTER TA
25 KIRAT LETTER THA
26 KIRAT LETTER DA
27 KIRAT LETTER DHA
28 KIRAT LETTER NA
29 Reserved
2A KIRAT LETTER PA
2B KIRAT LETTER PHA
2C KIRAT LETTER BA
2D KIRAT LETTER BHA
2E KIRAT LETTER MA
2F KIRAT LETTER YA
30 KIRAT LETTER RA
31 Reserved
32 KIRAT LETTER LA
33 Reserved
34 Reserved
35 KIRAT LETTER WA
36 KIRAT LETTER SHA
37 KIRAT LETTER SSA
38 KIRAT LETTER SA
39 KIRAT LETTER HA

3A Reserved
3B Reserved
3C Reserved
3D KIRAT VOWEL SIGN AH (tit-cha)
3E KIRAT VOWEL SIGN AA
3F KIRAT VOWEL SIGN I
40 KIRAT VOWEL SIGN II
41 KIRAT VOWEL SIGN U
42 KIRAT VOWEL SIGN UU
43 Reserved
44 Reserved
45 Reserved
46 KIRAT VOWEL SIGN AE (peh-cha)
47 KIRAT VOWEL SIGN E
48 KIRAT VOWEL SIGN AI
49 Reserved
4A Reserved
4B KIRAT VOWEL SIGN O
4C KIRAT VOWEL SIGN AU
4D KIRAT VIRAMA (sa-i)
4E Reserved
4F KIRAT LENGTH MARK (kehmphreng)
50 Reserved
51 Reserved
52 Reserved
53 Reserved
54 Reserved
55 Reserved
56 Reserved
57 Reserved
58 Reserved
59 Reserved
5A KIRAT STOP
5B Reserved
5C KIRAT SUBSCRIPT YA
5D KIRAT SUBSCRIPT RA
5E KIRAT SUBSCRIPT WA
5F Reserved

65 KIRAT CONJUNCT TR
66 KIRAT RENDERING FORM OF JA
67 KIRAT FINAL CONSONANT K
68 KIRAT FINAL CONSONANT NG
69 Reserved
6A KIRAT FINAL CONSONANT T
6B KIRAT FINAL CONSONANT N
6C KIRAT FINAL CONSONANT P
6D KIRAT FINAL CONSONANT M
6E KIRAT FINAL CONSONANT R
6F KIRAT FINAL CONSONANT L



Manipuri (Meithei)

The Manipuri script is a recently extinct script that was formerly used to write the Meithei language in Manipur State, India. The script may have been introduced as early as the fourteenth century CE or as late as the sixteenth. The only available source has been Grierson (see below).

The script is of the same lineage as Devanagari. Unlike Devanagari, there are no independent signs for vowels other than *a*, the other independent vowels being expressed as signs upon the independent vowel *a* (similar to the Tibetan method). The consonantal and vowel systems are both fairly complete, so it is probably most useful and correct to encode it in the ISCII manner, parallel to Devanagari as much as possible.

The *anusvara* (nasalization) mark in Manipuri produces some special rendering forms depending on the vowel preceding it. There are eight of these, producing the endings *ang, -ng, -ng, -ing, -eng, -ung, ng,* and *-ong*. The rendering forms look like ligatures of the vowel sign with the *anusvara*, or similar. Manipuri contains no long O vowel, so the place of the long O is filled with the diphthong sign AO, which does not seem to fit elsewhere.

Issues: Because Manipuri lacks special symbols for the independent vowels, the entire first column of an encoding completely parallel to Devanagari would be empty but for *anusvara* and the letter A. Therefore, to save one column, these have been moved into the column containing the consonants, so that A occurs just before KA, and the *anusvara* is left in the third position of that same row. The script can thus be put into four rows instead of five. There are presumably digits belonging to Manipuri, but no samples have been available. Space for them is available in the fifth column of the chart. It is also not known how much scholarly and historical interest there is in the Manipuri script.

Grierson, G. A. *Linguistic Survey of India,* Vol. 3, pt. 3., Bombay?, 1898?

Manipuri Names draft, mostly parallel to ISCII, 1992-10-23

```
00
01
02 MANIPURI ANUSVARA
03
04 MANIPURI LETTER A
05 MANIPURI LETTER KA
06 MANIPURI LETTER KHA
07 MANIPURI LETTER GA
08 MANIPURI LETTER GHA
09 MANIPURI LETTER NGA
0A MANIPURI LETTER CA
0B MANIPURI LETTER CHA
0C MANIPURI LETTER JA
0D MANIPURI LETTER JHA
0E MANIPURI LETTER NYA
0F MANIPURI LETTER TTA

10 MANIPURI LETTER TTHA
11 MANIPURI LETTER DDA
12 MANIPURI LETTER DDHA
13 MANIPURI LETTER NNA
14 MANIPURI LETTER TA
15 MANIPURI LETTER THA
16 MANIPURI LETTER DA
17 MANIPURI LETTER DHA
18 MANIPURI LETTER NA
```

```
19
1A MANIPURI LETTER PA
1B MANIPURI LETTER PHA
1C MANIPURI LETTER BA
1D MANIPURI LETTER BHA
1E MANIPURI LETTER MA
1F MANIPURI LETTER YA

20 MANIPURI LETTER RA
21
22 MANIPURI LETTER LA
23
24
25 MANIPURI LETTER WA
26 MANIPURI LETTER SHA
27 MANIPURI LETTER SSA
28 MANIPURI LETTER SA
29 MANIPURI LETTER HA
2A MANIPURI LETTER KSHA
2B
2C
2D
2E MANIPURI VOWEL SIGN AA
2F MANIPURI VOWEL SIGN I

30 MANIPURI VOWEL SIGN II
31 MANIPURI VOWEL SIGN U
32 MANIPURI VOWEL SIGN UU
33
34
35
36 MANIPURI VOWEL SIGN E
37
38 MANIPURI VOWEL SIGN AI
39 MANIPURI VOWEL SIGN OI
3A MANIPURI VOWEL SIGN O
3B MANIPURI VOWEL SIGN OI
3C MANIPURI VOWEL SIGN AU
3D MANIPURI VIRAMA
3E
3F

40 MANIPURI DIGIT ZERO
41 MANIPURI DIGIT ONE
42 MANIPURI DIGIT TWO
43 MANIPURI DIGIT THREE
44 MANIPURI DIGIT FOUR
45 MANIPURI DIGIT FIVE
46 MANIPURI DIGIT SIX
47 MANIPURI DIGIT SEVEN
48 MANIPURI DIGIT EIGHT
49 MANIPURI DIGIT NINE
4A
4B
4C
4D
4E
4F
```

Tai Lü (Chieng Mai, Northern Thai)

The Tai Lü script is widely used for various Tai dialects in northern Thailand, Yunnan, and parts of Myanmar (they are variously referred to as Lannathai, Yuan, or Kam Muang). The Tai Lü script is of the Brahmic variety, and is structurally similar to both the Thai and Myanmar scripts to which the affinities can be easily seen in the forms. The script is also known by the name *Northern Thai*; neither name seems to be a standard. The script referred to as *Chieng Mai* by Nakanishi is a fancier typographical form of the Tai Lü script, and hence included here. The *language* known as Tai Lü is in use in northern Thailand and in Yunnan province of China. There are about 1 million speakers of Tai Lu, and this script is officially recognized by the Chinese government.

Each Tai Lü consonant has an inherent vowel and (apparently) an inherent tone. Most of the consonants contain an inherent “o” vowel (or “a”?), but some seem to contain other inherent vowels. There are 41 consonants, five stand-alone vowels, and 32 vowel signs. The vowel system of the Northern Thai language is very complex, so the script contains a correspondingly large number of vowel signs, though some of them are written as compounds of simpler graphic symbols.

The traditional order of the consonants as given by Davis is distinctly different from the typical Devanagari order (for instance, the *aspirated* letters all come before the associated unaspirated ones, while Devanagari order is the opposite).

Issues: This draft is nowhere near complete as not enough is known at this time and sources are currently scarce. The chart is thought to contain a complete repertoire of possible candidates for encoding, except for punctuation and digits.

The vowel system could be greatly reduced by removing several compound vowel signs and manufacturing these vowels from simpler vowels and glyphic fragments. The glottal stop consonant itself is a component of the graphic representation of two other vowel signs.

The letters at codepoints 1B, 1D, 1E, 1F may be conjuncts of some type involving 18 together with other letters. Perhaps: MA=1B=18+13, LA=1D=18+14, NYA=1E=18+07, NGA=1F=18+03.

The names list is fully inadequate for any purpose except unique identification. The names were generated by taking Davis's pseudo-IPA transliterations and formulating unique names from them, while utilizing only the symbols allowed in ISO names.

Because the order cited by Davis differs so significantly from the Devanagari order, the utility and correctness of this order should be corroborated by other sources.

Davis, Richard. *A Northern Thai Reader.*
Pontalis, Pierre Lefevre. *L'invasion Thaie en Indo-Chine.*

Tai Lü (Chieng Mai, Northern Thai) names, rev 1992-10-21

```
00 TAI LU LETTER KHA
01 TAI LU LETTER KA
02 TAI LU LETTER KHAA1
03 TAI LU LETTER NGAA
04 TAI LU LETTER SA1
05 TAI LU LETTER CAA
06 TAI LU LETTER SAA1
07 TAI LU LETTER NYAA
08 TAI LU LETTER LAATHA
09 TAI LU LETTER LAADA
```

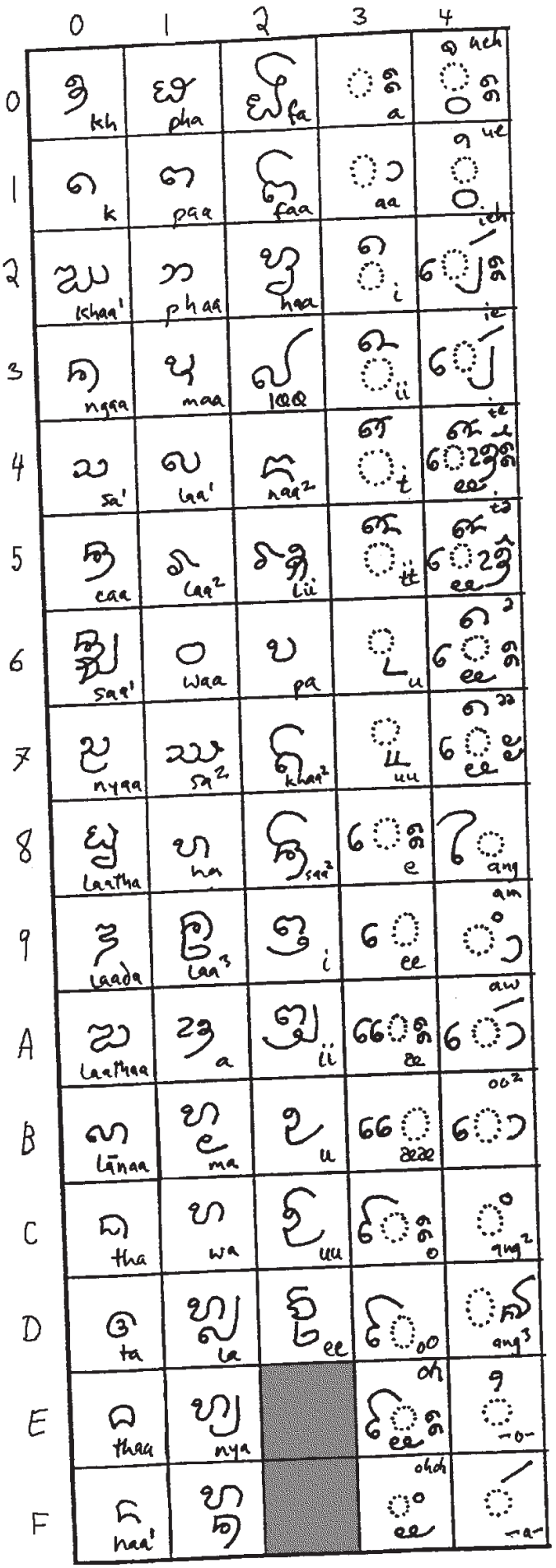
0A TAI LU LETTER LAATHAA
0B TAI LU LETTER LAANAA
0C TAI LU LETTER THA
0D TAI LU LETTER TAA
0E TAI LU LETTER THAA
0F TAI LU LETTER NAA1

10 TAI LU LETTER PHA
11 TAI LU LETTER PAA
12 TAI LU LETTER PHAA
13 TAI LU LETTER MAA
14 TAI LU LETTER LAA1
15 TAI LU LETTER LAA2
16 TAI LU LETTER WAA
17 TAI LU LETTER SA2
18 TAI LU LETTER HA
19 TAI LU LETTER LAA3
1A TAI LU LETTER A
1B TAI LU LETTER MA
1C TAI LU LETTER WA
1D TAI LU LETTER LA
1E TAI LU LETTER NYA
1F TAI LU LETTER NGA

20 TAI LU LETTER FA
21 TAI LU LETTER FAA
22 TAI LU LETTER HAA
23 TAI LU LETTER LAEAE
24 TAI LU LETTER NAA2
25 TAI LU LETTER LII
26 TAI LU LETTER PA
27 TAI LU LETTER KHAA2
28 TAI LU LETTER SAA2
29 TAI LU LETTER I
2A TAI LU LETTER II
2B TAI LU LETTER U
2C TAI LU LETTER UU
2D TAI LU LETTER EE
2E
2F

30 TAI LU VOWEL SIGN A
31 TAI LU VOWEL SIGN AA
32 TAI LU VOWEL SIGN I
33 TAI LU VOWEL SIGN II
34 TAI LU VOWEL SIGN I BAR
35 TAI LU VOWEL SIGN II BAR
36 TAI LU VOWEL SIGN U
37 TAI LU VOWEL SIGN UU
38 TAI LU VOWEL SIGN E
39 TAI LU VOWEL SIGN EE
3A TAI LU VOWEL SIGN AE
3B TAI LU VOWEL SIGN AEAE
3C TAI LU VOWEL SIGN O
3D TAI LU VOWEL SIGN OO
3E TAI LU VOWEL SIGN OH
3F TAI LU VOWEL SIGN OHOH

40 TAI LU VOWEL SIGN UEH
41 TAI LU VOWEL SIGN UE



42 TAI LU VOWEL SIGN IEH
43 TAI LU VOWEL SIGN IE
44 TAI LU VOWEL SIGN I BAR E
45 TAI LU VOWEL SIGN I BAR SCHWA
46 TAI LU VOWEL SIGN SCHWA
47 TAI LU VOWEL SIGN SCHWA SCHWA
48 TAI LU VOWEL SIGN ANG
49 TAI LU VOWEL SIGN AM
4A TAI LU VOWEL SIGN AW
4B TAI LU VOWEL SIGN OO TWO
4C TAI LU VOWEL SIGN ANG TWO
4D TAI LU VOWEL SIGN ANG THREE
4E TAI LU VOWEL SIGN O MEDIAL
4F TAI LU VOWEL SIGN A MEDIAL