Unicode Technical Report #3

Exploratory Proposals

The Unicode Consortium

Unicode Technical Report #3

Exploratory Proposals

The name **Unicode™** is a trademark of Unicode, Inc.

For the sake of brevity throughout this text
the word 'Unicode' alone is used to refer to
the Unicode Standard.

Unicode Technical Report #3
Exploratory Proposals

Review period closes August 15, 1993
Another draft will subsequently be issued for review

Introduction

This Technical Report is comprised of several exploratory proposals that the Unicode Technical Committee wishes to present for their first public review and commentary. These proposals have been generated from the committee's current knowledge about the scripts in question. Most of them are believed to be reasonable technical solutions for encoding of particular scripts, as far as can be ascertained at this time. However, many of them are known to be incomplete or be possessed of significant unresolved issues. The major unresolved issues are discussed in each proposal.

Technical inaccuracies and ambiguities are to be expected in a work of this nature, and most probably abound in these proposals. The work involves conjecture, relies on scanty information, and often requires re-interpretation as new information becomes available. The committee is *not* strongly committed to these proposals as they stand, and further information is being actively sought. Suggestions for improvement by way of additional symbols, further technical requirements, changes in the script model, refinements to the block introductions, or any other information can be mailed to the Scripts Subcommittee at the Unicode, Inc. address. The committee especially wishes to invite active participation and feedback from the communities which these proposals are designed to serve.

In these exploratory proposals, it is often mentioned that "sufficient information is not available" for some particular aspect of the script under discussion. This does not refer to the availability of information in an absolute sense, rather that the committee has not yet been able to obtain sufficient information for its archives.

Acknowledgements

Many individuals, too numerous to list here, have contributed information over a period of over a year during which portions of this report have been in preparation. The Unicode Technical Committee wishes to thank them collectively for their contributions, and hopes to see more such involvement in the future.

- The Glagolitic proposal was written by Joe Becker.
- All other proposals herein were written by Rick McGowan.

The following individuals have made significant contributions of time and energy in following bibliographic leads, searching libraries, forwarding information for the archives, or in analysis of various scripts included here:

Scott DeLancey, Lloyd Anderson, Andy Daniels, Elizabeth McGowan, Joan Aliprand, Glenn Adams, Lars-Erik Fredriksson, Asmus Freytag

About the Epigraphic Blocks

Semitic Alphabets

In these exploratory proposals, we distinguish two major "Early Semitic Alphabet" blocks, **Phoenician** and **Early Aramaic**, which are divided based on what may be termed "significant" differences in the shapes of various letters. Admittedly, this is a highly subjective choice. This arrangement makes two decisive cuts in a historical continuum covering several thousand years of middle-eastern history. The first cut is at approximately the point where several scripts leading eventually to the Aramaic and Hebrew branches began to be quite differentiated in their appearances from the branch that led to Punic. The second cut is at the point where the Aramaic/Hebrew branch began to noticeably split apart into the various lines that led to the Greek, Etruscan, and Latin branches on the one hand, and the Syriac, Arabic, and Hebrew branches on the other.

The alphabet encoded in the **Early Phoenician** block represents Phoenician as it stabilized by about 1100-1050 BC, as well as several early scripts that are quite closely related, though they are used to write a number of languages. The Phoenician block may be used, with appropriate font changes, to express Early Phoenician, Moabite, Early Hebrew, the earliest Early Aramaic, and Canaanite or Proto-Sinaitic scripts. It is also recommended for use to express Later Phoenician and **Punic**, which represent the main line of Phoenician evolution as a distinct script.

Later Branches of the Phoenician Alphabet

For encoding of **Late Aramaic** (especially papyri), **Palmyrene**, and **Nabataean** the Early Aramaic block should be used. The dividing line is relatively fuzzy, but in general a decision of which block to use can be made on the language, or when necessary on the general appearance of the script. The Unicode blocks are based rather roughly on "significant" differences in at least 12 letters (out of 22), including most obviously the letters transcribed as *A(aleph), B, H-underdot, T-underdot, Y, S, and R.* (A reasonable comparative source chart is contained in Healey's *The Early Alphabet*, fig. 15; the two blocks are divided approximately between the fourth and fifth of eight columns.)

Related Historical Script Blocks

- **South Arabian** and its descendents used for the **Lihyanite, Safaitic,** and **Thamudic** languages are encoded in the South Arabian block.
- The **Syriac** scripts (**Serta, Estrangela,** and **Nestorian** and their immediate precursors such as **Mandaic**) are encoded in a Syriac block and treated as font differences from a prototypical Syriac script. (Mandaic shapes are also shown in the Syriac block.) Varieties of Syriac are in modern use.
- **Etruscan** and **Oscan** are encoded in the Etruscan block.

Scripts Not Considered for Encoding

- **Lydian, Lycian , Sidetic, Carian** are not currently being considered for encoding. Information on the repertoire for the first two is available, but other significant information is lacking for all of them. They may eventually be encoded separately, or mapped onto other scripts.

Future Directions

In the future, this epigraphic introduction may be expanded to include further discussions of epigraphic scripts and families of scripts.

Some Sources

Healey, John F. *The Early Alphabet.*
Cross, Frank Moore. *The Invention and Development of the Alphabet.*
Encyclopaedia Brittanica, Articles on: *Anatolian Languages, Ancient Epigraphic Remains, Alphabets, Luwian, Lycian alphabet, Lycian language, Lydian language.*

Rev 92/11/25

0	𐡀	𐡃						
1	𐡁	𐡄						
2	𐡂	𐡅						
3	𐡃	𐡆						
4	𐡄	𐡇						
5	𐡅	𐡈						
6	𐡆							
7	𐡇							
8	𐡈							
9	𐡉							
A	𐡊							
B	𐡋							
C	𐡌							
D	𐡍							
E	𐡎							
F	𐡏							

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 for further information and issues.

Some Sources

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

Rev 92/10/30

Aramaic Names List, draft 92/10/29

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

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

0								
1								
2								
3								
4								
5								
6								
7								
8								
9								
A								
B								
C								
D								
E								
F								

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 when the people converted to Islam. It is related to the Arabic script.

In contrast to many other Indic 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 Indic 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 Indic scripts, or to Arabic. Not enough information is available at this time to determine the completeness of the accompanying chart. The digits 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.

Some Sources

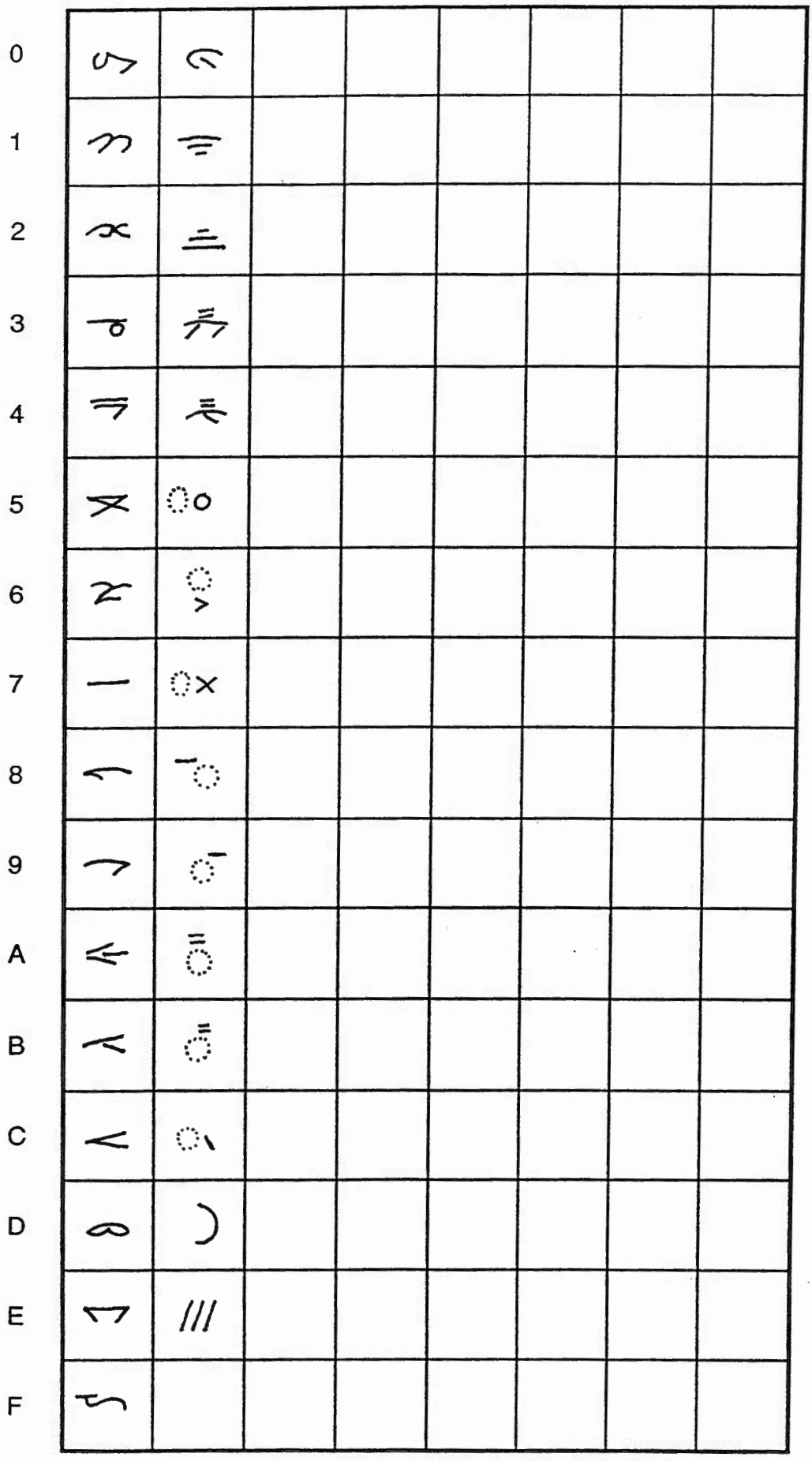
Grierson, G. A. *Linguistic Survey of India*, Vol. 3.
One photocopy of 2 pages (326 and 327) from an unknown volume in German.

Rev 92/11/25

Balti Names, draft 92/10/23

00 BALTI LETTER A
01 BALTI LETTER B
02 BALTI LETTER P
03 BALTI LETTER T
04 BALTI LETTER G
05 BALTI LETTER HH
06 BALTI LETTER C
07 BALTI LETTER CH
08 BALTI LETTER D
09 BALTI LETTER R
0A BALTI LETTER Z
0B BALTI LETTER S
0C BALTI LETTER SH
0D BALTI LETTER K
0E BALTI LETTER L
0F BALTI LETTER M

10 BALTI LETTER N
11 BALTI LETTER H
12 BALTI LETTER J
13 BALTI LETTER KH
14 BALTI LETTER TH
15 BALTI LETTER TS
16 BALTI LETTER NG
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?



Batak

The Batak script is (or was) used to write Toba (or Toba-Batak), Mandailing, Dairi, and possibly other languages on the island of Sumatra . The alphabet is called *si-sija-sija* in Toba-Batak (van der Tuuk). Batak is read from left to right, but is often written similarly to Tagalog and Buhid, by writing vertically along the length of a piece of bamboo.

The phonetic system of the script is similar to the scripts of the Philippines (Tagalog). Like Tagalog and other scripts of the archipelagos between Southeast Asia and Australia, Batak ultimately derives from scripts of India. Batak has a *virama* and final consonants *are* expressed in the script. Like Tagalog, only two independent vowels other than *a* are included in the script (but several vowel signs are used). The alphabetical order (if van der Tuuk gives it in order) differs from both the primeval Sanskritic and Tagalog orders; the accompanying chart is in the order given for Toba-Batak.

The vowel signs *i, o,* and the *pangolat* (=*virama*) are spacing marks. The vowel signs *e* and final *ng* are non-spacing marks. The vowel sign *i* is placed after the consonant. The vowel sign *u* is placed under a consonant and somewhat to the right. Several ligated forms of letters with the *u* sound are known. The vowel sign *o* is placed after the consonant. The pangolet is likewise placed after the consonant, causing the inherent *a* vowel to be lost. The *final ng* is placed above the consonant and somewhat to the right. (When *e* and *ng* occur together on a consonant, thus, there are two dashlike marks above.) The *hamisaran* is usually written above the vowels *i* and *o*. When *pangolat* (the devoweller) is used to close a syllable, the vowel sign for the *previous* vowel is placed either under the final consonant or after the final consonant, and before the *pangolat* itself.

Punctuation is not normally used, all letters simply running together, but a *bindu* does exist and is occasionally used to disambiguate similar words or phrases. (This *bindu* is unfortunately known by the same name as the *virama, pangolat.*) The *bindu* apparently appears in several forms. One is called *bindu pinardjolma* and is used to separate sections of text; another is *bindu pinarulok*, and a third is *bindu pinarboras*, again used to separate sections of text. These marks are apparently large signs that physically separate sections of text, and may be more in the manner of ornaments than characters. Thus, only one *bindu* mark is included in the chart. A sign called *pustaha* is also sometimes used to separate a

title from the main text which normally begins on the same line.

Mandailing: The Mandailing alphabetical order differs somewhat from Toba-Batak, and North Mandailing again differs slightly from South Mandailing. Some of the letter shapes are likewise slightly different; these are *ha* and *sa*. The rendering forms for the consonant vowel-sign combinations *pa+u, sa+u,* and *la+u* may differ from the forms used for Toba Batak. Mandailing uses two other letters for *k* and *tj* sounds. These two letters are produced by putting a mark called *tompi* onto the normal letters for *h* and *s*. It is not known whether the *tompi* is otherwise productive, so both the Mandailing letters and the *tompi* itself are included in the chart.

Dairi: Dairi alphabetical order again differs from Toba-Batak and Mandailing. Dairi does not include the letter *nja*. The forms for *ta* and *wa* differ significantly from those used for Toba-Batak. The vowel sign listed in the chart as *u* is pronounced more like a closed *e* and written after the associated consonant rather than under (or attached to) the consonant. The sign *sikordjan*, which is pronounced as a soft *h* following the associated vowel, is placed over the consonant. When *final ng* is used in Dairi, it goes over the previous consonant rather than over the vowel sign. In Toba-Batak, it may optionally go over the vowel if the vowel is not a non-spacing mark.

Issues: It is not clear whether the Mandailing *tompi* is different from the Dairi *sikordjan*; if not, then one of them should be deleted from the chart.

Batak is known to have been in use in the mid-1800s. Nakanishi (1975) states that it is "seldom used today." It may be extinct as of this writing (1992). The completeness of this analysis and chart is not known.

Some Sources

van der Tuuk, H. N. *A Grammar of Toba Batak.*

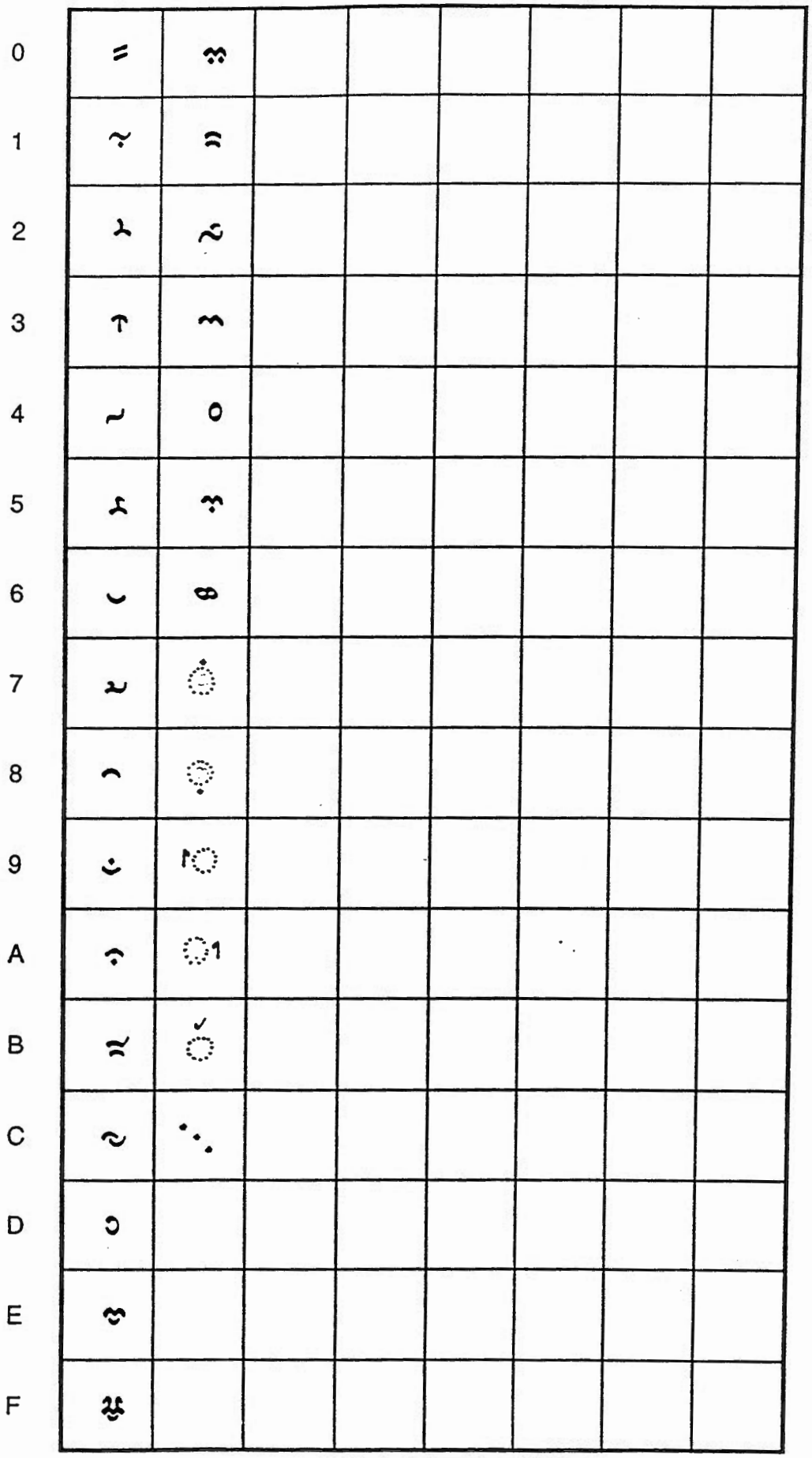
Rev 92/10/23

Batak Names draft, 92/10/23

00 BATAK LETTER A
01 BATAK LETTER HA
02 BATAK LETTER MA
03 BATAK LETTER NA
04 BATAK LETTER RA
05 BATAK LETTER TA
06 BATAK LETTER SA
07 BATAK LETTER PA
08 BATAK LETTER LA
09 BATAK LETTER GA
0A BATAK LETTER DJA
0B BATAK LETTER DA
0C BATAK LETTER NGA
0D BATAK LETTER BA
0E BATAK LETTER WA
0F BATAK LETTER JA

10 BATAK LETTER NJA
11 BATAK LETTER I
12 BATAK LETTER U
13 MANDAILING LETTER K
14 MANDAILING LETTER TJ
15 BATAK VOWEL SIGN I (HALUAIN)
16 BATAK VOWEL SIGN U (HABORUWAN)
17 BATAK VOWEL SIGN O (SIJALA)
18 BATAK VOWEL SIGN E (HATADINGAN)
19 BATAK FINAL NG (HAMISARAN)
1A MANDAILING DIACRITICAL MARK TOMPI
1B DAIRI SOFT H SIGN SIKORDJAN
1C BATAK VIRAMA PANGOLAT
1D BATAK SEPARATOR (BINDU)
1E BATAK SIGN PUSTAHA

BUGINESE 92/10/29



Buginese

The Buginese script is used on the island of Sulawesi, mainly in the south-west. It is of the Indic type and perhaps related to Javanese. It bears some affinity with Tagalog as well, and it apparently does not record final consonants. Buginese may be the easternmost representative of the Brahmi descendents. Sirk (1983) reports that the Buginese language (an Austronesian language) has a rich traditional literature making it one of the foremost languages of Indonesia. There may be as many as 2.3 million speakers of Buginese in the southern part of Sulawesi (as of 1971). The script was reported in some use as of 1983, and a variety of traditional literature has been printed in it.

Buginese literature was studied extensively by B. F. Matthes (a Dutch missionary) in the 19th century. Matthes published a Buginese-Dutch dictionary in 1874 with a supplement in 1889, as well as a grammar.. The script was previously also used to write the Makassarese, Bimanese, and Madurese languages.

Buginese seems to use spaces between certain units, which are noted by Sirk to be "longer than a word in its grammatical definition." There is one punctuation symbol, *pallawa*, used "to separate rhythmico-intonational groups, thus functionally corresponding to the full stop and comma of the Latin script." It is also apparently used sometimes to denote word doubling.

Issues: The only page from Fossey available to this author (page 377) comments that the ordering, also observed here, is after Matthes, and further remarks on "*une certaine différence entre les caractères de ses publications et ceux de l'Imprimerie Nationale.*" The digits, if any, are unknown.

Some Sources

Nakanishi, Akira. *Writing Systems of the World.*
Fossey, Charles. Notices sur les caract©ères ©étrangers, an
Sirk, Ü. *The Buginese Language.*

Rev 92/11/25

Buginese Names draft, 92/10/23

00 BUGINESE LETTER KA
01 BUGINESE LETTER GA
02 BUGINESE LETTER NNA
03 BUGINESE LETTER NNKA
04 BUGINESE LETTER PA
05 BUGINESE LETTER BA
06 BUGINESE LETTER MA
07 BUGINESE LETTER MPA
08 BUGINESE LETTER TA
09 BUGINESE LETTER DA
0A BUGINESE LETTER NA
0B BUGINESE LETTER NRA
0C BUGINESE LETTER CA
0D BUGINESE LETTER JA
0E BUGINESE LETTER NYA
0F BUGINESE LETTER NYCA

00 BUGINESE LETTER YA
11 BUGINESE LETTER RA
12 BUGINESE LETTER LA
13 BUGINESE LETTER WA
14 BUGINESE LETTER SA
15 BUGINESE LETTER A
16 BUGINESE LETTER HA
17 BUGINESE VOWEL SIGN I
18 BUGINESE VOWEL SIGN U
19 BUGINESE VOWEL SIGN E ACUTE
1A BUGINESE VOWEL SIGN O
1B BUGINESE VOWEL SIGN E BREVE
1C BUGINESE PUNCTUATION MARK

0	Ꭰ	Ꮀ	Ꮐ	Ꮠ	Ꮰ	Ᏸ		
1	Ꭱ	Ꮁ	Ꮑ	Ꮡ	Ꮱ	Ᏹ		
2	Ꭲ	Ꮂ	Ꮒ	Ꮢ	Ꮲ	Ᏺ		
3	Ꭳ	Ꮃ	Ꮓ	Ꮣ	Ꮳ	Ᏻ		
4	Ꭴ	Ꮄ	Ꮔ	Ꮤ	Ꮴ	Ᏼ		
5	Ꭵ	Ꮅ	Ꮕ	Ꮥ	Ꮵ			
6	Ꭶ	Ꮆ	Ꮖ	Ꮦ	Ꮶ			
7	Ꭷ	Ꮇ	Ꮗ	Ꮧ	Ꮷ			
8	Ꭸ	Ꮈ	Ꮘ	Ꮨ	Ꮸ			
9	Ꭹ	Ꮉ	Ꮙ	Ꮩ	Ꮹ			
A	Ꭺ	Ꮊ	Ꮚ	Ꮪ	Ꮺ			
B	Ꭻ	Ꮋ	Ꮛ	Ꮫ	Ꮻ			
C	Ꭼ	Ꮌ	Ꮜ	Ꮬ	Ꮼ			
D	Ꭽ	Ꮍ	Ꮝ	Ꮭ	Ꮽ			
E	Ꭾ	Ꮎ	Ꮞ	Ꮮ	Ꮾ			
F	Ꭿ	Ꮏ	Ꮟ	Ꮯ	Ꮿ			

Cherokee Syllabary

The Cherokee script is a syllabic system used by the Cherokee Indians of North America. It was invented in the early 19th Century by Sequoyah who, realizing the power of written language, set out to produce a system of writing for his language. It was first tested among the Western Cherokee, and quickly adopted by the tribal council. The modern syllabary consists of 85 letters. There actually exist two forms of each letter; the modern symbols (shown here) are apparently the result of the need for simplified forms to be used with 19th century typesetting technology. As originally invented, the symbols were all much more cursive in form (see the sample in Alexander's *Dictionary*).

Modern Cherokee punctuation and page formatting conventions are as in English. Though the Cherokee syllabary is caseless, capitalization has been observed in some publications for proper names and at the beginning of each sentence, however, the "majuscule" letters do not differ at all in appearance from the minuscule letters, they are merely of larger size. Though Sequoyah invented a system of numerals for Cherokee, they were not adopted by the tribal council and have never been used. There are thus no independent digits encoded in the Cherokee block; Arabic (Western) digits are used.

Encoding Structure: The Unicode block for the Cherokee script is arranged in linear order consistent with what seems to be its normal collation order. The columnar arrangement below is the typical arrangement shown in dictionaries and textbooks. The vowel written as "v" is a nasalized "u" (after Holmes & Smith). No syllable *mv* exists.

Syllabary Layout

A		E	I	O	U	V
GA	KA	GE	GI	GO	GU	GV
HA		HE	HI	HO	HU	HV
LA		LE	LI	LO	LU	LV
MA		ME	MI	MO	MU	--
NA	HNA NAH	NE	NI	NO	NU	NV
QUA		QUE	QUI	QUO	QUU	QUV
SA	S	SE	SI	SO	SU	SV
DA	TA	DE TE	DI TI	DO	DU	DV
DLA	TLA	TLE	TLI	TLO	TLU	TLV
TSA		TSE	TSI	TSO	TSU	TSV
WA		WE	WI	WO	WU	WV
YA		YE	YI	YO	YU	YV

Other Issues: It may be advisable to include an 86th symbol, which was invented but quickly fell out of use. It occurs in facsimiles of pages in Sequoyah's hand. Its phonetic value has been reported as being close to that of HV.

Some Sources

Holmes, Ruth Bradley and Betty Sharp Smith. *Beginning Cherokee.*
Alexander, J. T. *A Dictionary of the Cherokee Indian Language.*
Sloat, Clarence, et al. *Introduction to Phonology.*
Kilpatrick, Jack Frederick and Anna Gritts Kilpatrick, eds. *New Echota Letters.*

Rev 92/10/29

Draft Cherokee Names List, 10/20/92.

00 CHEROKEE LETTER A
01 CHEROKEE LETTER E
02 CHEROKEE LETTER I
03 CHEROKEE LETTER O
04 CHEROKEE LETTER U
05 CHEROKEE LETTER V
06 CHEROKEE LETTER GA
07 CHEROKEE LETTER KA
08 CHEROKEE LETTER GE
09 CHEROKEE LETTER GI
0A CHEROKEE LETTER GO
0B CHEROKEE LETTER GU
0C CHEROKEE LETTER GV
0D CHEROKEE LETTER HA
0E CHEROKEE LETTER HE
0F CHEROKEE LETTER HI

10 CHEROKEE LETTER HO
11 CHEROKEE LETTER HU
12 CHEROKEE LETTER HV
13 CHEROKEE LETTER LA
14 CHEROKEE LETTER LE
15 CHEROKEE LETTER LI
16 CHEROKEE LETTER LO
17 CHEROKEE LETTER LU
18 CHEROKEE LETTER LV
19 CHEROKEE LETTER MA
1A CHEROKEE LETTER ME
1B CHEROKEE LETTER MI
1C CHEROKEE LETTER MO
1D CHEROKEE LETTER MU
1E CHEROKEE LETTER NA
1F CHEROKEE LETTER HNA

20 CHEROKEE LETTER NAH
21 CHEROKEE LETTER NE
22 CHEROKEE LETTER NI
23 CHEROKEE LETTER NO
24 CHEROKEE LETTER NU
25 CHEROKEE LETTER NV
26 CHEROKEE LETTER QUA
27 CHEROKEE LETTER QUE
28 CHEROKEE LETTER QUI
29 CHEROKEE LETTER QUO
2A CHEROKEE LETTER QUU

2B CHEROKEE LETTER QUV
2C CHEROKEE LETTER SA
2D CHEROKEE LETTER S
2E CHEROKEE LETTER SE
2F CHEROKEE LETTER SI

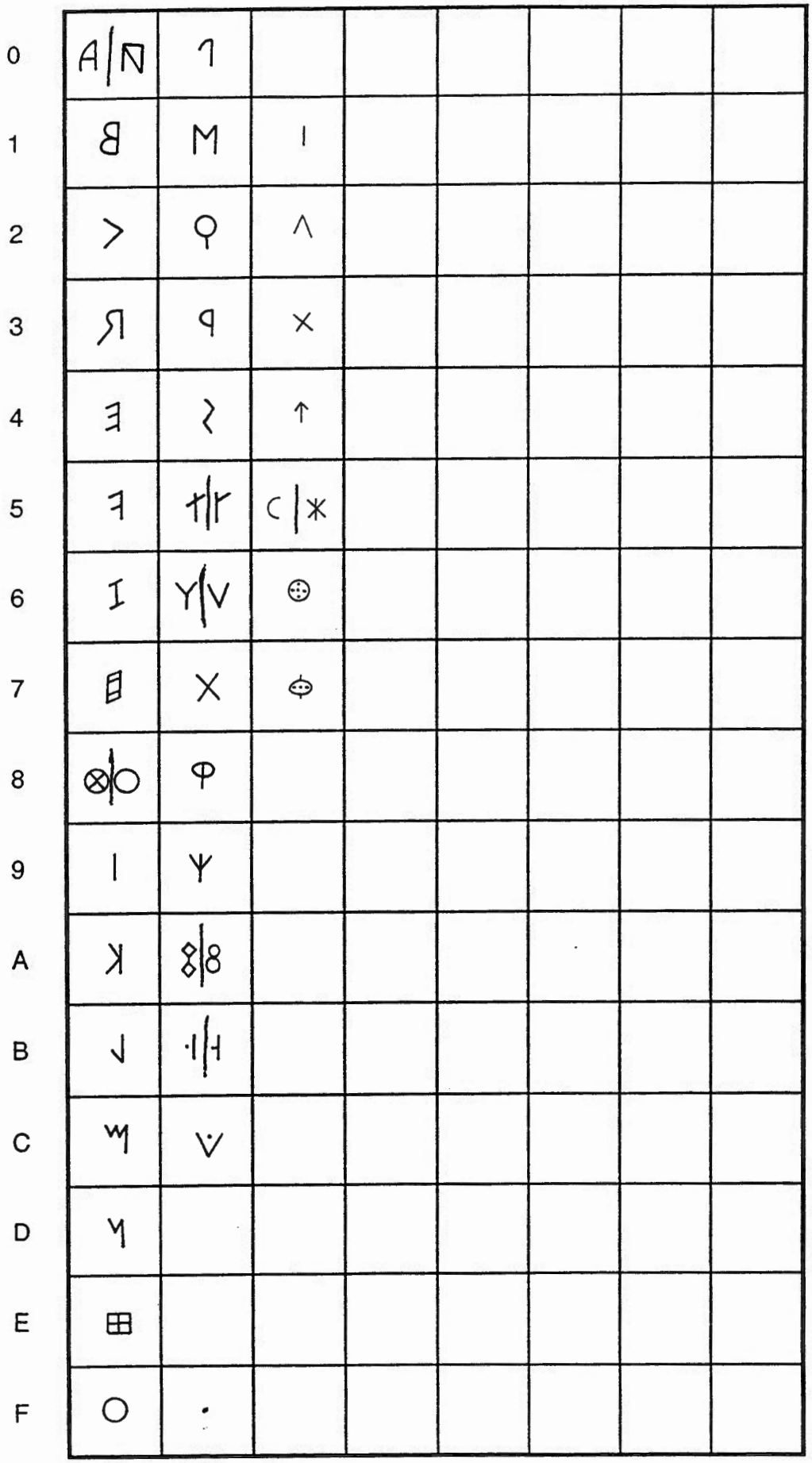
30 CHEROKEE LETTER SO
31 CHEROKEE LETTER SU
32 CHEROKEE LETTER SV
33 CHEROKEE LETTER DA
34 CHEROKEE LETTER TA
35 CHEROKEE LETTER DE
36 CHEROKEE LETTER TE
37 CHEROKEE LETTER DI
38 CHEROKEE LETTER TI
39 CHEROKEE LETTER DO
3A CHEROKEE LETTER DU
3B CHEROKEE LETTER DV
3C CHEROKEE LETTER DLA
3D CHEROKEE LETTER TLA
3E CHEROKEE LETTER TLE
3F CHEROKEE LETTER TLI

40 CHEROKEE LETTER TLO
41 CHEROKEE LETTER TLU
42 CHEROKEE LETTER TLV
43 CHEROKEE LETTER TSA
44 CHEROKEE LETTER TSE
45 CHEROKEE LETTER TSI
46 CHEROKEE LETTER TSO
47 CHEROKEE LETTER TSU
48 CHEROKEE LETTER TSV
49 CHEROKEE LETTER WA
4A CHEROKEE LETTER WE
4B CHEROKEE LETTER WI
4C CHEROKEE LETTER WO
4D CHEROKEE LETTER WU
4E CHEROKEE LETTER WV
4F CHEROKEE LETTER YA

50 CHEROKEE LETTER YE
51 CHEROKEE LETTER YI
52 CHEROKEE LETTER YO
53 CHEROKEE LETTER YU
54 CHEROKEE LETTER YV

Etruscan

draft
DEC 10, 1991



Etruscan

The Etruscan script is used to write both the Etruscan and Oscan (or Oscan-Umbrian) languages. Etruscan was the language of a people (who called themselves *rasna*) in Etruria, corresponding to modern Tuscany in western Italy. The Etruscan civilization lived alongside the Romans and there was much contact between the two. Inscriptions in Etruscan date from about the 7th century BC through the first century AD. The Etruscan and Oscan languages are unrelated, Oscan being an Italic language similar to Latin and Etruscan being imperfectly known and of uncertain linguistic affiliation.

Etruscan is written horizontally from right to left (occasionally boustrophedon). Archaic inscriptions have no spaces between words, but later inscriptions frequently have single or double dots between words. The letters *ii* and *uu* are used in Oscan but not in Etruscan. The letters *s* and *o* (0E and 0F) appear in Etruscan inscriptions *only* in the context of abecedaries and were apparently not used in writing the Etruscan language.

Etruscan numerals are imperfectly known. They are similar to Roman numerals, but they are read and written *from right to left*, in contrast to Latin. The numerals at 26 and 27 are uncertain.

Issues: The numerals are too uncertain at this time to warrant a final encoding; more information is necessary.

Some Sources

Encyclopaedia Brittanica, Article: *Etruscan Language*.
Bonfante, Larissa. *Etruscan*.

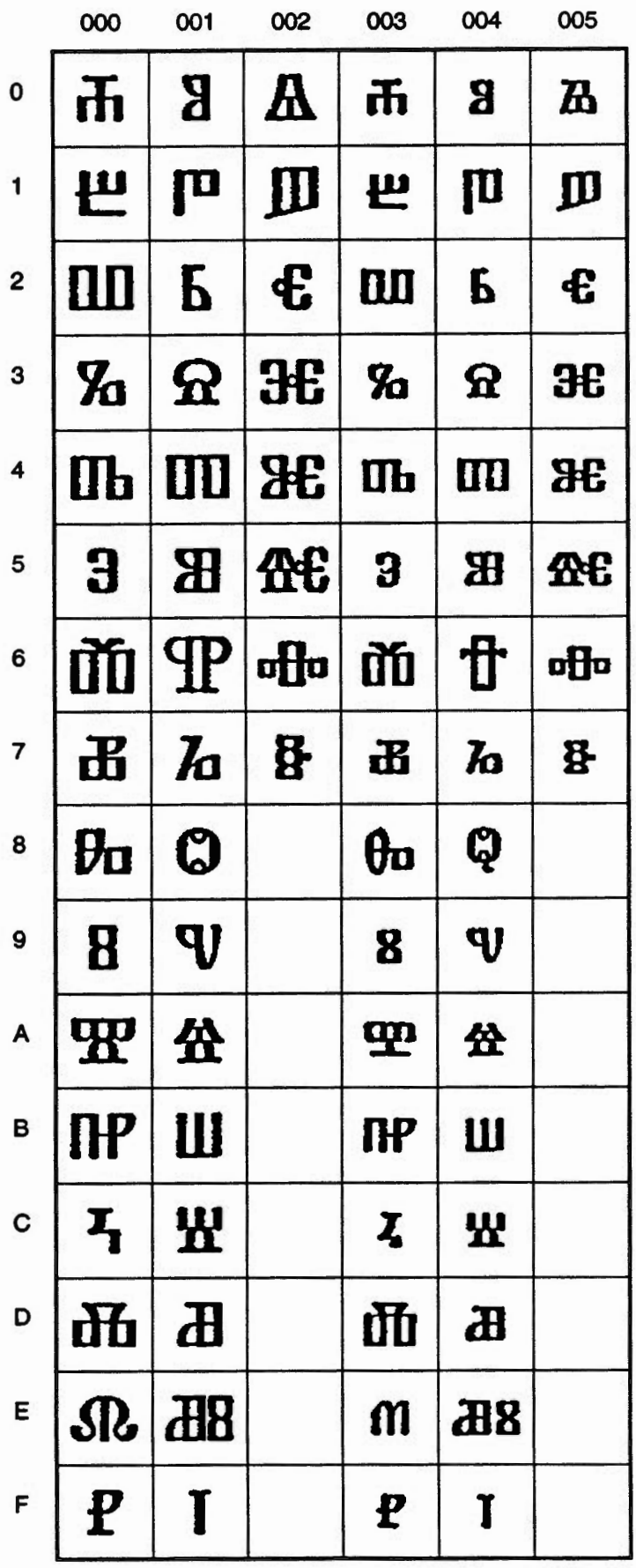
Rev 92/10/20

Etruscan Names List, draft 92/10/29

00 ETRUSCAN LETTER A
01 ETRUSCAN LETTER B
02 ETRUSCAN LETTER C
03 ETRUSCAN LETTER D
04 ETRUSCAN LETTER E
05 ETRUSCAN LETTER V
06 ETRUSCAN LETTER Z
07 ETRUSCAN LETTER H
08 ETRUSCAN LETTER TH
09 ETRUSCAN LETTER I
0A ETRUSCAN LETTER K
0B ETRUSCAN LETTER L
0C ETRUSCAN LETTER M
0D ETRUSCAN LETTER N
0E ETRUSCAN LETTER S
0F ETRUSCAN LETTER O

10 ETRUSCAN LETTER P
11 ETRUSCAN LETTER SH
12 ETRUSCAN LETTER Q
13 ETRUSCAN LETTER R
14 ETRUSCAN LETTER S
15 ETRUSCAN LETTER T
16 ETRUSCAN LETTER U
17 ETRUSCAN LETTER SS
18 ETRUSCAN LETTER PH
19 ETRUSCAN LETTER KH
1A ETRUSCAN LETTER F
1B ETRUSCAN LETTER II
1C ETRUSCAN LETTER UU
1D
1E
1F

20
21 ETRUSCAN NUMERAL I
22 ETRUSCAN NUMERAL V
23 ETRUSCAN NUMERAL X
24 ETRUSCAN NUMERAL L
25 ETRUSCAN NUMERAL C
26 ETRUSCAN NUMERAL UNKNOWN A
27 ETRUSCAN NUMERAL UNKNOWN B


000 001 002 003 004 005
0 1 2 3 4 5 6 7 8 9 A B C D E F

Glagolitic

Glagolitic, sometimes called by its Russian name *Glagolitsa* (''verbal script''), was developed in the 9th century to write Old Slavic. It arose more or less in parallel with the Cyrillic alphabet for the same language, and the two alphabets correspond to each other quite closely. The relationship between the origins of Glagolitic and Cyrillic is unknown, though St. Cyril is said to have had a hand in both. The Cyrillic script gradually supplanted Glagolitic, but Glagolitic continued in some liturgical use until the 19th century.

In the encoding, Glagolitic is treated as a separate script from Cyrillic, principally because the letter shapes are in most cases totally unrelated, with differences not at all arising from "mere font style". Glagolitic itself is seen in two slightly different styles, called the Bulgarian-Macedonian and Croatian. The Croatian form distinguishes uppercase and lowercase letters, although the difference in nearly all instances is merely one of size. The letterforms shown in the charts are Croatian style.

Like Cyrillic, the Glagolitic script is written in linear sequence from left to right with no contextual modification of the letterforms.

Variant Glyph Forms: Two or three of the letters have variant glyph forms. These are not given separate character codes.

Encoding Order: The ordering is basically the same as that of the (old) Cyrillic alphabet. Occasional sources show minor variations in the ordering of one or two characters.

Letter Names: These old names for the Cyrillic letters apply as well to the Glagolitic.

Encoding Structure: The Unicode block for the Glagolitic script is divided into the following ranges:

U+00	to	U+27	Uppercase letters (generic Glagolitic)
U+28	to	U+2F	Currently unassigned
U+30	to	U+57	Lowercase letters (Croatian-style only)
U+58	to	U+5F	Currently unassigned

Open issues:

1. order and names of IZHE / I: seems to be random, may be able to find a preference.

2. discrepancies with (DIS) 6861

- it appears to contain 3 pairs of variant glyphs for the same letters
  - suggest ignoring these, there's room to add them later if necessary
- it appears to contain 1 (or 2) pairs of letters seen nowhere else
  - suggest ignoring these, there's room to add them later if appropriate
- it appears to contain 1 duplicated glyph (IZHE)
  - suggest ignoring this, apparently a mistake

DRAFT GLAGOLITIC CHARACTER NAMES LIST

0 Uppercase letters (generic Glagolitic)
00 GLAGOLITIC CAPITAL LETTER AZ
01 GLAGOLITIC CAPITAL LETTER BUKI
02 GLAGOLITIC CAPITAL LETTER VEDI
03 GLAGOLITIC CAPITAL LETTER GLAGOL
04 GLAGOLITIC CAPITAL LETTER DOBRO
05 GLAGOLITIC CAPITAL LETTER YEST
06 GLAGOLITIC CAPITAL LETTER ZHIVETE
07 GLAGOLITIC CAPITAL LETTER ZELO
08 GLAGOLITIC CAPITAL LETTER ZEMLYA
09 GLAGOLITIC CAPITAL LETTER IZHE
0A GLAGOLITIC CAPITAL LETTER I
= izhey
0B GLAGOLITIC CAPITAL LETTER DERV
= gerv
0C GLAGOLITIC CAPITAL LETTER KAKO
0D GLAGOLITIC CAPITAL LETTER LYUDI
0E GLAGOLITIC CAPITAL LETTER MISLETE
0F GLAGOLITIC CAPITAL LETTER NASH
10 GLAGOLITIC CAPITAL LETTER ON
11 GLAGOLITIC CAPITAL LETTER POKOY
12 GLAGOLITIC CAPITAL LETTER RTSI
13 GLAGOLITIC CAPITAL LETTER SLOVO
14 GLAGOLITIC CAPITAL LETTER TVERDO
15 GLAGOLITIC CAPITAL LETTER UK
16 GLAGOLITIC CAPITAL LETTER FERT
17 GLAGOLITIC CAPITAL LETTER KHER
18 GLAGOLITIC CAPITAL LETTER OT
= omega
19 GLAGOLITIC CAPITAL LETTER TSI
1A GLAGOLITIC CAPITAL LETTER CHERV
1B GLAGOLITIC CAPITAL LETTER SHA
1C GLAGOLITIC CAPITAL LETTER SHTA
1D GLAGOLITIC CAPITAL LETTER YER
1E GLAGOLITIC CAPITAL LETTER YERI
1F GLAGOLITIC CAPITAL LETTER YERY
20 GLAGOLITIC CAPITAL LETTER YAT
21 GLAGOLITIC CAPITAL LETTER YU
22 GLAGOLITIC CAPITAL LETTER YUS MALIY
23 GLAGOLITIC CAPITAL LETTER YUS MALIY YOTIROVANNIY
24 GLAGOLITIC CAPITAL LETTER YUS BOLSHOY
25 GLAGOLITIC CAPITAL LETTER YUS BOLSHOY YOTIROVANNIY
26 GLAGOLITIC CAPITAL LETTER FITA
27 GLAGOLITIC CAPITAL LETTER IZHITSA
28

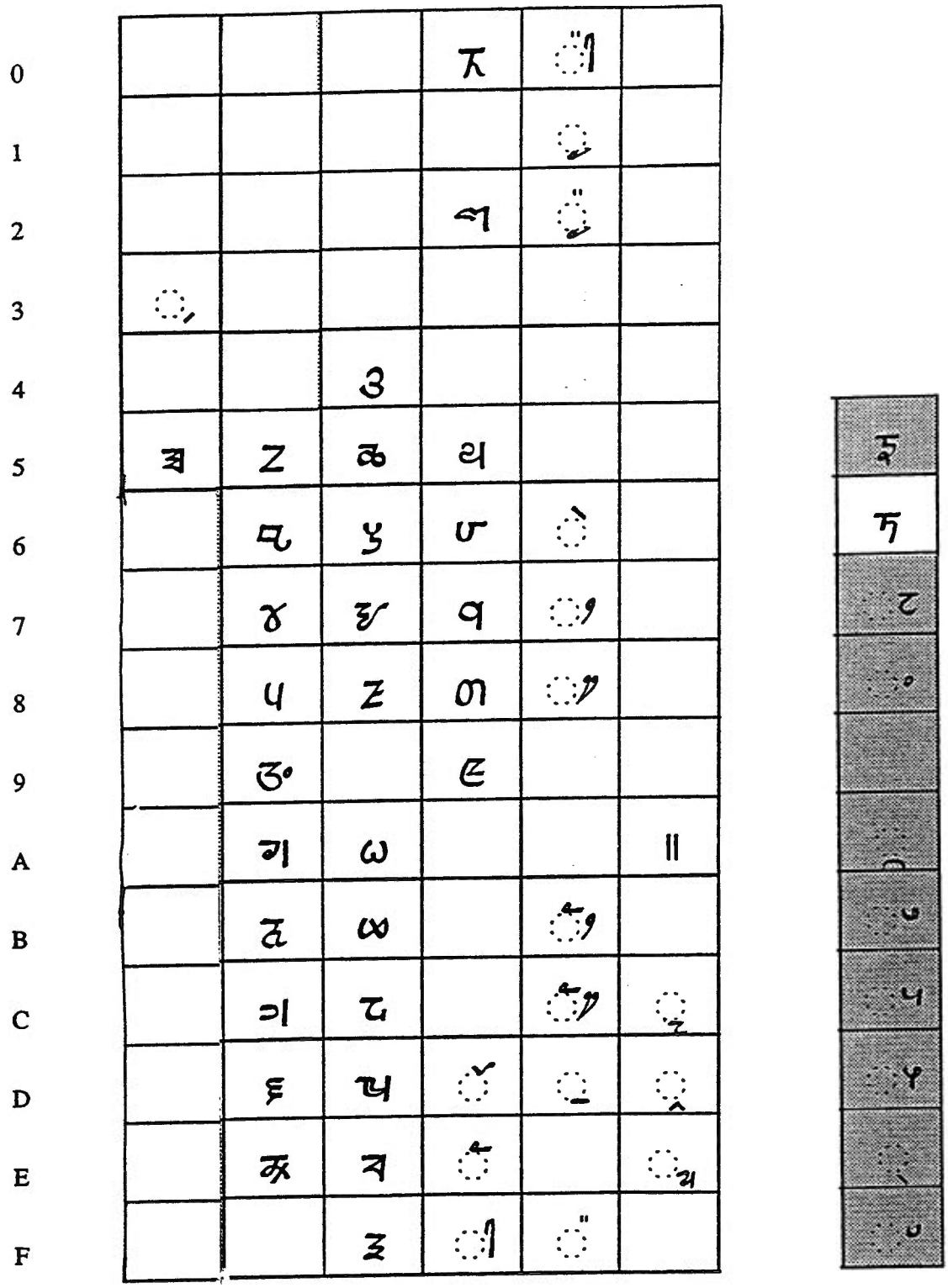
29
2A
2B
2C
2D
2E
2F

@ Lowercase letters (Croatian-style only)
30 GLAGOLITIC SMALL LETTER AZ
31 GLAGOLITIC SMALL LETTER BUKI
32 GLAGOLITIC SMALL LETTER VEDI
33 GLAGOLITIC SMALL LETTER GLAGOL
34 GLAGOLITIC SMALL LETTER DOBRO
35 GLAGOLITIC SMALL LETTER YEST
36 GLAGOLITIC SMALL LETTER ZHIVETE
37 GLAGOLITIC SMALL LETTER ZELO
38 GLAGOLITIC SMALL LETTER ZEMLYA
39 GLAGOLITIC SMALL LETTER IZHE
3A GLAGOLITIC SMALL LETTER I
= izhey
3B GLAGOLITIC SMALL LETTER DERV
= gerv
3C GLAGOLITIC SMALL LETTER KAKO
3D GLAGOLITIC SMALL LETTER LYUDI
3E GLAGOLITIC SMALL LETTER MISLETE
3F GLAGOLITIC SMALL LETTER NASH
40 GLAGOLITIC SMALL LETTER ON
41 GLAGOLITIC SMALL LETTER POKOY
42 GLAGOLITIC SMALL LETTER RTSI
43 GLAGOLITIC SMALL LETTER SLOVO
44 GLAGOLITIC SMALL LETTER TVERDO
45 GLAGOLITIC SMALL LETTER UK
46 GLAGOLITIC SMALL LETTER FERT
47 GLAGOLITIC SMALL LETTER KHER
48 GLAGOLITIC SMALL LETTER OT
= omega
49 GLAGOLITIC SMALL LETTER TSI
4A GLAGOLITIC SMALL LETTER CHERV
4B GLAGOLITIC SMALL LETTER SHA
4C GLAGOLITIC SMALL LETTER SHTA
4D GLAGOLITIC SMALL LETTER YER
4E GLAGOLITIC SMALL LETTER YERI
4F GLAGOLITIC SMALL LETTER YERY
50 GLAGOLITIC SMALL LETTER YAT
51 GLAGOLITIC SMALL LETTER YU
52 GLAGOLITIC SMALL LETTER YUS MALIY

53 GLAGOLITIC SMALL LETTER YUS MALIY YOTIROVANNIY
54 GLAGOLITIC SMALL LETTER YUS BOLSHOY
55 GLAGOLITIC SMALL LETTER YUS BOLSHOY YOTIROVANNIY
56 GLAGOLITIC SMALL LETTER FITA
57 GLAGOLITIC SMALL LETTER IZHITSA
58
59
5A
5B
5C
5D
5E
5F

KIRAT / LIMBU

92/10/30
(92/2/2)



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 Róng (Lepcha) script. It has 20 consonants (including the stand-alone "A" as in other Indic 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 Róng 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 therefor 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 "jñ" 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). Because of these observations, the final consonants should, perhaps be encoded distinctly and no invisible *virama* encoded. In this case, Limbu would then be similar to the model used for Róng. See also the block introduction for Róng (Lepcha).

In either case, the script bears some similarity to the Róng 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.

Some Sources

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.

Rev 92/10/30

Kirat (Limbu) Names List, draft 92/10/20
This is a sign inventory of the chart rather than a names list. The chart follows the ISCII order, as discussed in the Issues section of the block introduction; the names for each codepoint may be obtained by looking at the Unicode Devanagari block.

KIRAT LETTER KA
KIRAT LETTER KHA
KIRAT LETTER GA
KIRAT LETTER NGA
KIRAT LETTER CHA
KIRAT LETTER CHHA
KIRAT LETTER JA
KIRAT LETTER NA
KIRAT LETTER TA
KIRAT LETTER THA
KIRAT LETTER DA
KIRAT LETTER DHA
KIRAT LETTER PA
KIRAT LETTER PHA
KIRAT LETTER BA
KIRAT LETTER BHA
KIRAT LETTER MA
KIRAT LETTER YA
KIRAT LETTER RA
KIRAT LETTER LA
KIRAT LETTER WA
KIRAT LETTER SHA
KIRAT LETTER SA
KIRAT LETTER HA
KIRAT LETTER GHA
KIRAT LETTER A

KIRAT VOWEL SIGN A
KIRAT VOWEL SIGN I
KIRAT VOWEL SIGN U
KIRAT VOWEL SIGN E
KIRAT VOWEL SIGN AI
KIRAT VOWEL SIGN O
KIRAT VOWEL SIGN AU
KIRAT VOWEL SIGN TIT-CHA
KIRAT VOWEL SIGN PET-CHA

KIRAT FINAL CONSONANT K
KIRAT FINAL CONSONANT NG
KIRAT FINAL CONSONANT T
KIRAT FINAL CONSONANT N

KIRAT FINAL CONSONANT P
KIRAT FINAL CONSONANT M
KIRAT FINAL CONSONANT R
KIRAT FINAL CONSONANT L

KIRAT SUBSCRIPT YA
KIRAT SUBSCRIPT RA
KIRAT SUBSCRIPT WA

KIRAT ASPIRATION MARK (MUKPHRENG)
KIRAT LENGTH MARK (KEHMPHRENG)
KIRAT VIRAMA? (SAI)

KIRAT ANUSVARA
KIRAT PROLONGED FINAL MARK
KIRAT STOP

Linear B

The script called *Linear B* is a syllabic system that was used on the island of Crete (and parts of the nearby mainland) to write the oldest recorded variety of the Greek language. Linear B clay tablets predate Homeric Greek by some 700 years, the latest being from about 1375 BC. Major archaeological sites include Knossos, first uncovered in about 1900 by Sir Arthur Evans, and a major site near Pylos on the mainland. The majority of inscriptions currently known are inventories of commodities and accounting records.

The script resisted early attempts at decipherment, but it finally yielded to the efforts of Michael Ventris, an architect and amateur decipherer. Ventris' breakthrough in decipherment came after the realization that the language might be Greek, and not (as had been previously thought) a completely unknown language. Ventris formed an alliance with John Chadwick, and decipherment proceeded quickly. Ventris and Chadwick published a joint paper in 1953.

Linear B was written from left to right with no non-spacing marks or other complications. The script consists mainly of a number of phonetic signs representing the combination of a consonant and vowel. There are 60 known phonetic signs, a few signs that seem to be mainly free variants (Chadwick's *optional signs*), a few unidentified signs, numerals, and a number of ideographic signs which were used mainly as counters for commodities. Some ligatures formed from combinations of syllables were apparently used as well. Chadwick gives several examples of these ligatures, which are not included in this encoding.

The signs having phonetic values beginning with *J* are pronounced in the German manner as the English *Y*.

Issues: The first four rows (through the syllable *zo*) are well established; the rest of the symbols are more questionable. Some of the unknown symbols may now be known, and hence require some movement of codes. The characters for weights are not necessarily in a sensible order. There may be no distinction between characters 43 and 6A. The ideograms (e.g., for weight) may be the tip of a much larger ideographic iceberg, though the sources would seem to indicate that there are only a small number of such ideograms. The 5th unknown symbol may be *gold*, but it's not clear; one older source listed it as unknown, but Chadwick's book (see below) lists it as meaning *gold*. The character names for the weight units

reflect the lists in Chadwick, but *do not* convey the proper meaning well; better names must be found.

The historical importance of Linear B is well established. It may make sense, however, to encode Linear B along with **Linear A** and the **Cypriot Syllabary** of Enkomi, either as a unified set of signs or separately in adjacent blocks with phonetic parallels. Unicode archives contain some references for Linear A and Cypriot.

The Linear B ligatures may be another case requiring the encoding of some form of ligature manufacturing code in Unicode, since such ligatures would be optional and totally free variants in any rendering system. Such a ligature code has been widely discussed, and may be necessary in other scripts as well.

Some Sources

Chadwick, John. *Linear B and Related Scripts.*
Sampson, Geoffrey. *Writing Systems; a linguistic introduction.*

Rev 92/11/25

Linear B names, 92/10/26

00 LINEAR B SYLLABLE A
01 LINEAR B SYLLABLE E
02 LINEAR B SYLLABLE I
03 LINEAR B SYLLABLE O
04 LINEAR B SYLLABLE U
05 LINEAR B SYLLABLE DA
06 LINEAR B SYLLABLE DE
07 LINEAR B SYLLABLE DI
08 LINEAR B SYLLABLE DO
09 LINEAR B SYLLABLE DU
0A LINEAR B SYLLABLE JA
0B LINEAR B SYLLABLE JE
0C
0D LINEAR B SYLLABLE JO
0E LINEAR B SYLLABLE JU
0F LINEAR B SYLLABLE KA

10 LINEAR B SYLLABLE KE
11 LINEAR B SYLLABLE KI
12 LINEAR B SYLLABLE KO
13 LINEAR B SYLLABLE KU
14 LINEAR B SYLLABLE MA
15 LINEAR B SYLLABLE ME
16 LINEAR B SYLLABLE MI
17 LINEAR B SYLLABLE MO
18 LINEAR B SYLLABLE MU (OX)
19 LINEAR B SYLLABLE NA
1A LINEAR B SYLLABLE NE
1B LINEAR B SYLLABLE NI (FIGS)
1C LINEAR B SYLLABLE NO
1D LINEAR B SYLLABLE NU
1E LINEAR B SYLLABLE PA
1F LINEAR B SYLLABLE PE

20 LINEAR B SYLLABLE PI
21 LINEAR B SYLLABLE PO
22 LINEAR B SYLLABLE PU
23 LINEAR B SYLLABLE QA
24 LINEAR B SYLLABLE QE
25 LINEAR B SYLLABLE QI (SHEEP)
26 LINEAR B SYLLABLE QO
27
28 LINEAR B SYLLABLE RA
29 LINEAR B SYLLABLE RE
2A LINEAR B SYLLABLE RI

2B LINEAR B SYLLABLE RO
2C LINEAR B SYLLABLE RU
2D LINEAR B SYLLABLE SA (FLAX)
2E LINEAR B SYLLABLE SE
2F LINEAR B SYLLABLE SI

30 LINEAR B SYLLABLE SO
31 LINEAR B SYLLABLE SU
32 LINEAR B SYLLABLE TA
33 LINEAR B SYLLABLE TE
34 LINEAR B SYLLABLE TI
35 LINEAR B SYLLABLE TO
36 LINEAR B SYLLABLE TU
37 LINEAR B SYLLABLE WA
38 LINEAR B SYLLABLE WE
39 LINEAR B SYLLABLE WI
3A LINEAR B SYLLABLE WO
3B
3C LINEAR B SYLLABLE ZA
3D LINEAR B SYLLABLE ZE
3E
3F LINEAR B SYLLABLE ZO

40
41 LINEAR B SYLLABLE HA
42 LINEAR B SYLLABLE INITIAL AI
43 LINEAR B SYLLABLE INITIAL AU
44 LINEAR B SYLLABLE DWE
45 LINEAR B SYLLABLE DWO
46 LINEAR B SYLLABLE NWA
47 LINEAR B SYLLABLE PA3
48 LINEAR B SYLLABLE PHU
49 LINEAR B SYLLABLE PTE
4A LINEAR B SYLLABLE RJA
4B LINEAR B SYLLABLE RAI (SAFFRON)
4C LINEAR B SYLLABLE RJO
4D LINEAR B SYLLABLE SWA
4E LINEAR B SYLLABLE SWI
4F LINEAR B SYLLABLE TJA

50 LINEAR B SYLLABLE TWO
51 LINEAR B UNKNOWN SYMBOL 1
52 LINEAR B UNKNOWN SYMBOL 2
53 LINEAR B UNKNOWN SYMBOL 3
54 LINEAR B UNKNOWN SYMBOL 4
55 LINEAR B UNKNOWN SYMBOL 5
56 LINEAR B UNKNOWN SYMBOL 6

57 LINEAR B UNKNOWN SYMBOL 7
58 LINEAR B UNKNOWN SYMBOL 8
59 LINEAR B UNKNOWN SYMBOL 9
5A LINEAR B UNKNOWN SYMBOL 10
5B LINEAR B SYLLABLE TWE
5C LINEAR B IDEOGRAM CLOTH
5D LINEAR B IDEOGRAM WHEAT
5E LINEAR B IDEOGRAM WINE
5F LINEAR B IDEOGRAM BRONZE

60 LINEAR B IDEOGRAM WOOL
61 LINEAR B IDEOGRAM BARLEY
62 LINEAR B IDEOGRAM OLIVE OIL
63 LINEAR B IDEOGRAM GOLD
64 LINEAR B IDEOGRAM SHEEP
65 LINEAR B IDEOGRAM RAM
66 LINEAR B IDEOGRAM EWE
67 LINEAR B IDEOGRAM GOAT
68 LINEAR B IDEOGRAM HE-GOAT
69 LINEAR B IDEOGRAM SHE-GOAT
6A LINEAR B IDEOGRAM PIG
6B LINEAR B IDEOGRAM BOAR
6C LINEAR B IDEOGRAM SOW
6D LINEAR B IDEOGRAM OX
6E LINEAR B IDEOGRAM BULL
6F LINEAR B IDEOGRAM COW

70 LINEAR B WEIGHT TIMES SIX
71 LINEAR B WEIGHT TIMES TWELVE
72 LINEAR B WEIGHT TIMES FOUR
73 LINEAR B WEIGHT TIMES THIRTY
74 LINEAR B WEIGHT MAXIMUM
75 LINEAR B DRY WEIGHT TIMES FOUR
76 LINEAR B DRY WEIGHT TIMES SIX
77 LINEAR B DRY WEIGHT TIMES TEN
78 LINEAR B LIQUID MEASURE TIMES THREE

MALDIVIAN (Dihevi) DRAFT 10/25/91

	0	1	2					
0	ﷲ	ﷺ	ﷻ					
1	ﷳ	ﷴ	ﷵ					
2	ﷶ	ﷷ	ﷸ					
3	ﷹ	ﷺ	ﷻ					
4	﷼	﷽						
5	؀	؁	؂					
6	؃	؄	؅					
7	؆	؇	؈					
8	؉	؊	؋					
9	،؍	؎	؏					
A	ؐ	ؑ	ؒ					
B	ؓ	ؔ	ؕ					
C	ؖ	ؗ	ؘ					
D	ؙ	ؚ	؛					
E	؜	؝	؞					
F	؟	ؠ	ء					

Maldivian (Dihevi)

The Maldivian script is used in the Republic of Maldives (a group of atolls in the Indian Ocean, circa 400 miles SW of Sri Lanka, about 4°N 73°E) to write the *Dihevi* language.

Maldivian is written from right to left and partakes of features of both the Indic and Arabic script varieties. Consonants have an inherent *a* vowel sound, but they are always written with either a vowel sign or a null "vanishing vowel" sign (U+xx2A) above them. On *alif* (U+xx07) the null vowel sign is a glottal stop. Loanwords from Arabic are also written in the Arabic script or transcribed by means of dots on existing Maldivian letters. Both Arabic and Western digits are used.

Issues: There is also an older set of Maldivian letter forms (for which see Faulmann) which are completely different from, yet exactly parallels these. It should probably not be considered a separate script. The older form could be used by shifting fonts.

Encoding Structure: The Unicode block for the Maldivian script is divided into four ranges:

U+xx00 → U+xx17	Consonant Letters
U+xx18 → U+xx23	Extended Maldivian Letters
U+xx24	Currently unassigned
U+xx25 → U+xx2F	Non-spacing Vowel Signs

Issues: The enumeration of the 12 Extended Maldivian Letters used for transcriptions of Arabic letters is consistent with the Unicode treatment of the Arabic script, in which various combinations of dots are always alotted separate code points. The source of these is the Library of Congress *Cataloging Service Bulletin, No. 19 / Winter 1982*. The 12 text elements listed in that publication follow, in Arabic alphabetic order, with their Arabic equivalents:

Maldivian Character	Arabic Letter Equivalent
TH + triple overdot	THAA
H + underdot	HAA
H + overdot	KHAA
D + overdot	THAL
S + triple overdot	SHEEN
S + underdot	SAD

S	+	overdot	DAD
TH	+	underdot	TAH
TH	+	overdot	DHAH
A	+	underdot	AIN
A	+	overdot	GHAIN
G	+	double overdot	QAF

The idea that Maldivian letters have an inherent *a* vowel is from Nakanishi, but it seems inconsistent with the fact that the letters never appear without a vowel sign or a null-vowel sign. This issue must be clarified.

Some Sources

Nakanishi, Akira. *Writing Systems of the World.*
Library of Congress. *Cataloging Service Bulletin*, No. 19 / Winter 1982.
Faulmann, Carl. *Schriftzeichen und Alphabete aller Zeiten und Völker.*

Rev 92/11/25

Maldivian Names List, draft 92/10/29

(These names reflect only the phonetic values.)

00 MALDIVIAN LETTER H
01 MALDIVIAN LETTER SH
02 MALDIVIAN LETTER N
03 MALDIVIAN LETTER R
04 MALDIVIAN LETTER B
05 MALDIVIAN LETTER L
06 MALDIVIAN LETTER K
07 MALDIVIAN LETTER A
08 MALDIVIAN LETTER W,V
09 MALDIVIAN LETTER M
0A MALDIVIAN LETTER F,PH
0B MALDIVIAN LETTER D
0C MALDIVIAN LETTER TH
0D MALDIVIAN LETTER L
0E MALDIVIAN LETTER G
0F MALDIVIAN LETTER NY

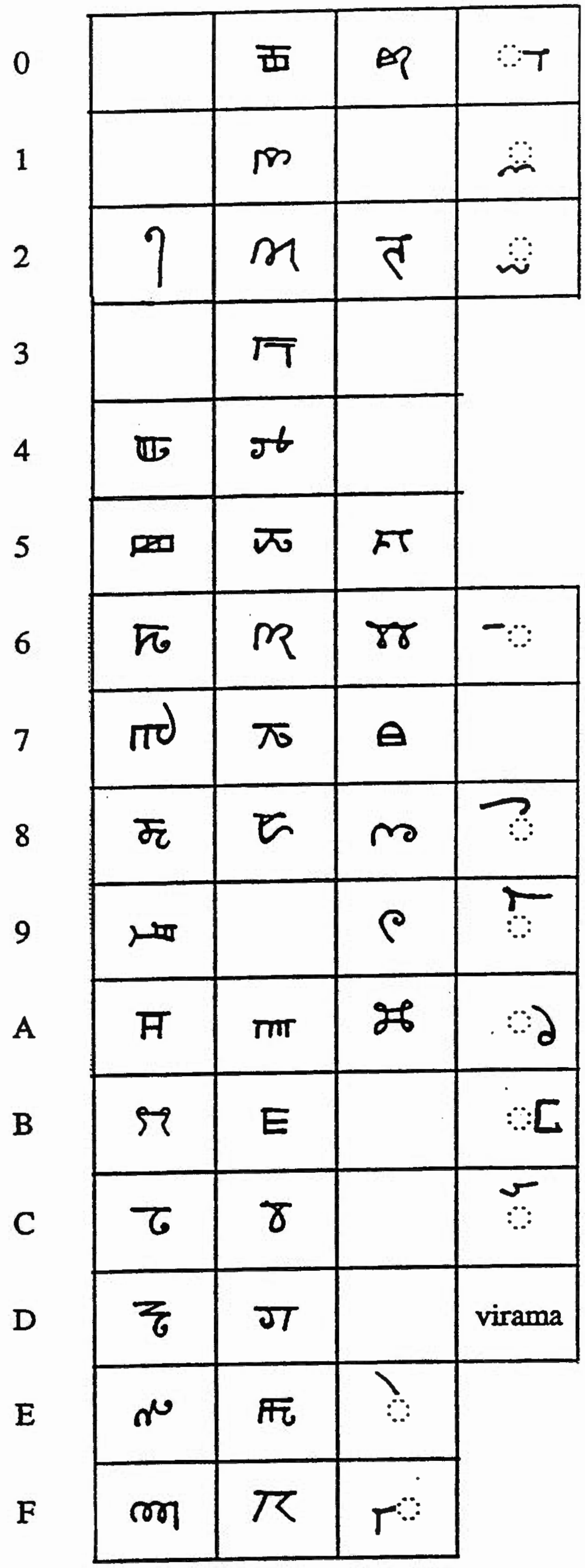
10 MALDIVIAN LETTER S
11 MALDIVIAN LETTER D
12 MALDIVIAN LETTER Z
13 MALDIVIAN LETTER T
14 MALDIVIAN LETTER Y
15 MALDIVIAN LETTER P
16 MALDIVIAN LETTER J
17 MALDIVIAN LETTER CH
18 MALDIVIAN LETTER TH WITH THREE DOTS ABOVE
19 MALDIVIAN LETTER H WITH DOT BELOW
1A MALDIVIAN LETTER H WITH DOT ABOVE
1B MALDIVIAN LETTER D WITH DOT ABOVE
1C MALDIVIAN LETTER S WITH THREE DOTS ABOVE
1D MALDIVIAN LETTER S WITH DOT BELOW
1E MALDIVIAN LETTER S WITH DOT ABOVE
1F MALDIVIAN LETTER TH WITH DOT BELOW

20 MALDIVIAN LETTER TH WITH DOT ABOVE
21 MALDIVIAN LETTER A WITH DOT BELOW
22 MALDIVIAN LETTER A WITH DOT ABOVE
23 MALDIVIAN LETTER G WITH two DOTS ABOVE
24
25 MALDIVIAN VOWEL SIGN A
26 MALDIVIAN VOWEL SIGN I
27 MALDIVIAN VOWEL SIGN U
28 MALDIVIAN VOWEL SIGN E

29 MALDIVIAN VOWEL SIGN O
2A MALDIVIAN VOWEL SIGN AA
2B MALDIVIAN VOWEL SIGN II
2C MALDIVIAN VOWEL SIGN UU
2D MALDIVIAN VOWEL SIGN EE
2E MALDIVIAN VOWEL SIGN OO
2F MALDIVIAN NULL VOWEL SIGN (Sukun)

MANIPURI

92/10/29
(92/2/2)



Manipuri (Meithei)

The Manipuri script is a recently extinct script that was formerly used to write the Methei language in Manipur State, India. The script may have been introduced as early as the fourteenth century 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*, *-áng*, *-íng*, *-ing*, *-eng*, *-ung*, *úng*, 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 dipthong 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.

Some Sources

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

Rev 92/11/25

Manipuri Names draft, mostly parallel to ISCII, 92/10/23

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

Meroitic

draft
DEC 10, 1991

	0	1					(X	Y)
0	𐦀	𐦁						
1	𐦂	𐦃						
2	𐦄	𐦅						
3	𐦆	𐦇						
4	𐦈	𐦉						
5	𐦊	𐦋						
6	𐦌	𐦍						
7	𐦎	𐦏						
8	𐦐							
9	𐦑							
A	𐦒							
B	𐦓							
C	𐦔							
D	𐦕							
E	𐦖							
F	𐦗							

Meroïtic

Meroïtic was the language of a great African kingdom (called *Kush*) which lay to the south of Egypt in what is now the Sudan. The capital city was Meroë (modern Begrawiya), along the Nile River. The Meroïtic script is a syllabary, and its glyphs are derived from or related to Egyptian Hieroglyphics. It comes in two forms, monumental (Hieroglyphic) and cursive, of which the monumental is much more rare. The two forms bear very little outward resemblance, the one looking very much like Egyptian, the other quite abbreviated, not unlike Demotic.

The earliest dated Meroïtic inscriptions are from about 180 BC, and it was extinct by the 5th Century AD. The Meroïtic script was first deciphered by F. L. Griffith in the early 1900s and that work was later refined somewhat by F. Hintze and others. The language itself, though, remains incompletely known in the absence of bilingual inscriptions and relationships to other known languages.

Most consonantal signs of Meroïtic have an inherent *a* vowel, except when they are followed by one of the vowel signs *i, e,* or *o*. There are special signs for the combinations *ne, se, te,* and *to*. Meroïtic is usually written from right to left in cursive form, and from top to bottom (with columns running from right to left) in monumental form. In the monumental form, the human and animal figures face *in the direction which the text runs* (i.e., away from the beginning of the line). It should be carefully noted that this is *unlike* Egyptian, in which the figures face the beginning of the line.

Issues: The main draft chart shows the cursive form, with corresponding hieroglyphic shapes in columns labelled X and Y. These have completely different values than identical Egyptian Hieroglyphic symbols, and unification of Meroïtic and Egyptian (if attempted) would be purely on the basis of glyphic identity in the monumental form, not on abstract letter semantics. Unification seems inadvisable because the normal form is the cursive form.

The ordering of symbols in the two main sources differs in the 3rd and 4th positions (*o* and *i*) and also in the 16th and 17th positions (*s* and *se*). The order used here is that given in *Friedrich*, while the transliteration is after *Davies*. There does not seem to be a standard order.

Some Sources

Davies, W. V. *Egyptian Hieroglyphs.*
Friedrich, Johannes. *Extinct Languages.*

Rev 92/10/21

Meroitic, draft Dec 10, 1991

00 MEROITIC LETTER A
01 MEROITIC LETTER E
02 MEROITIC LETTER O
03 MEROITIC LETTER I
04 MEROITIC LETTER Y
05 MEROITIC LETTER W
06 MEROITIC LETTER B
07 MEROITIC LETTER P
08 MEROITIC LETTER M
09 MEROITIC LETTER N
0A MEROITIC LETTER NE
0B MEROITIC LETTER R
0C MEROITIC LETTER L
0D MEROITIC LETTER H
0E MEROITIC LETTER HH
0F MEROITIC LETTER S

10 MEROITIC LETTER SE
11 MEROITIC LETTER K
12 MEROITIC LETTER Q
13 MEROITIC LETTER T
14 MEROITIC LETTER TE
15 MEROITIC LETTER TO
16 MEROITIC LETTER D
17 MEROITIC WORD DIVIDER

0								
1								
2								
3								
4								
5								
6								
7								
8								
9								
A								
B								
C								
D								
E								
F								

Ogham

The Ogham script was used in Ireland and England prior to the introduction of the Latin alphabet. The form of its letters seems heavily influenced by the medium with which it was used; it was most often scratched on stones and posts, as well as on the frames of doors. At least one interactive variety called "leg Ogham" (reported in the Book of Ballymote) was also apparently used; it was signed with the hands upon the shin, the five fingers being used in a manner suggesting the horizontal lines of the script.

The Ogham is divided into groups of five. The last five are diphthongs, and are later developments. Each letter has a traditional name which is the name of a tree or shrub. Some of the phonetic values apparently differ depending on the locale in which it was used and the language being written.

Ogham was formerly written on stones and door lintels from the bottom left hand side, over the crest, and down the right hand side. The center line in the charts represents the corner of a stone or lintel. It is suggested that it be rendered on computers from left to right, turned 90 degrees counterclockwise with the center line running horizontally, or top to bottom, with the center line running vertically.

Punctuation was not normally used in Ogham, but later developments suggest that a middle dot delimiter or a vertical line delimiter may be used; sources are unclear on this point.

Issues: There is distinct disagreement in the sources available as to the order of the first five letters. Ogham has been called "Beth-Luis-Nuin" possibly after the first three letters, but other sources say these are the first, second, and fifth letters. In either case, the sources thus give conflicting names for the latter three of the first five letters. This question must be resolved satisfactorily before a final encoding can be made. The present names are after Lehmann (see below).

Some Sources

Lehmann, Ruth P. M. *Ogham: Ancient Script of the Celts.*
Graves, Robert. *The White Goddess*

Rev 92/10/20

Ogham Draft Names List, 92/10/20

00 OGHAM LETTER BEITHE
01 OGHAM LETTER LUIS
02 OGHAM LETTER FERN
03 OGHAM LETTER SAIL
04 OGHAM LETTER NUIN

05 OGHAM LETTER HUATHE
06 OGHAM LETTER DUIR
07 OGHAM LETTER TINNE
08 OGHAM LETTER COLL
09 OGHAM LETTER CIERT

0A OGHAM LETTER MUINN
0B OGHAM LETTER GORT
0C OGHAM LETTER GETAL
0D OGHAM LETTER STRAIF
0E OGHAM LETTER RUIS

0F OGHAM LETTER AILM
10 OGHAM LETTER ONN
11 OGHAM LETTER UR
12 OGHAM LETTER EDAD
13 OGHAM LETTER IDAD

14 OGHAM LETTER EABAB
15 OGHAM LETTER OIR
16 OGHAM LETTER UILLEND
17 OGHAM LETTER IPHIN
18 OGHAM LETTER MO'R

0	𐬀	𐬁	𐬂					
1	𐬃	𐬄	𐬅					
2	𐬆	𐬇	𐬈					
3	𐬉	𐬊	𐬋					
4	𐬌	𐬍	𐬎					
5	𐬏	𐬐	𐬑					
6	𐬒	𐬓	𐬔					
7	𐬕 | 𐬖	𐬗	𐬘					
8	𐬙	𐬚	𐬛					
9	𐬜	𐬝	𐬞					
A	𐬟	𐬠	𐬡					
B	𐬢	𐬣	𐬤					
C	𐬥	𐬦	𐬧					
D	𐬨	𐬩	𐬪					
E	𐬫	𐬬						
F	𐬭	𐬮						

Pahlavi/Avestan

The Pahlavi script is an historically important script related to the Arabic script. It was used (in various related forms) over a period of nearly a thousand years to write Pazand, Middle Persian, Parthian, and Pahlavi languages. An improved form of Pahlavi which includes explicit vowel letters was used to write the Avesta (the sacred book of Zoroastrianism containing teachings of the prophet Zoroaster or Zarathushtra); the latter form of the script is referred to as Avestan.

Pahlavi is written from right to left, in the Arabic manner. The form known as Book Pahlavi contains only 13 simple letters, certain graphemes that originally represented distinct letters having been coalesced to a high degree. Avestan, on the other hand, is improved and the ambiguities are much less. The accompanying chart is intended for use with Pahlavi and Avestan both. The Avestan letter forms are shown, and some of the Book Pahlavi forms differ slightly from these.

Pahlavi utilizes a complex seemingly open-ended set of ligatures and pronounciation changes in various combinations. Many of the letters do some sort of "double duty." There are complex cursive connections between certain characters preceding or following. Some of the double-duty letters were sometimes written with diacritical marks or dots to remove ambiguities in some situations.

The Avestan alphabet, in contrast, is much more regular and the letters generally refer to a single phoneme. The set of vowel letters in Avestan is considerably improved, and there are fewer (or no) cursive connections. The letter called *ao* by Jackson is a ligature of *aa* + *schwa*.

Issues: The order given here is not very good. The main source for Avestan (Jackson) is mute regarding alphabetical order. There was a bit of detective work involved in generating correspondences between that and other sources on Book Pahlavi. The shapes in the accompanying chart are the Avestan shapes (after Jackson). The letter *aa* may be better unencoded, simply using *a* + *a*. A case could probably be made for having an abstract *length mark* which could be used for doubling the vowels. It seems to be the case that, except for *a*, the short vertical appendage below each vowel has the meaning of lengthening it.

Complete names for the Avestan letters being currently unavailable, the names list is a hodge-podge using a semblance of the phonetic value, mainly after Jackson. The numerals are not well specified in the sources available at this time; hence, no numerals are given in the accompanying chart.

Pahlavi seems to contain a large number of words called "ideograms" in the literature (see Nyberg, for instance) that appear to be words which are actually pronounced and have a meaning fairly unrelated to their "literal" meaning and pronounciation if viewed simply as a group of letters.

There are two important ligatures that stand for the endings *et, eh,* or *end.* None of the sources gave enough detail on the usage and etymology of these. It is also not clear whether some of the "letters" of Avestan given by Jackson should not be simple ligatures; these are *sk, s-ogonek-hacek, n-tilde, ao.* These are not shown in the accompanying chart.

Jackson seems to not give an alphabetical order. The Book Pahlavi alphabetical order should probably be followed, and this does that to some extent. However, the interpolation of some letters may mean that there are letters out of order here, and the order should be carefully considered.

Some Sources

Nyberg, Henrik Samuel. *A Manual of Pahlavi.*
Haug, Martin. *An Old Pahlavi-Pazand Glossary.*
Jackson, A. V. Williams. *An Avesta Grammar in Comparison with Sanskrit.*
MacKenzie, D. N. *A Concise Pahlavi Dictionary.*

Rev 92/10/30

Pahlavi Names, draft, 92/10/27

00 PAHLAVI LETTER A
01 PAHLAVI LETTER B
02 PAHLAVI LETTER P
03 PAHLAVI LETTER T
04 PAHLAVI AVESTAN LETTER T
05 PAHLAVI LETTER TH
06 PAHLAVI LETTER J
07 PAHLAVI LETTER CH
08 PAHLAVI LETTER KH
09 PAHLAVI LETTER D
0A PAHLAVI LETTER DH
0B PAHLAVI LETTER R
0C PAHLAVI LETTER Z
0D PAHLAVI LETTER S
0E PAHLAVI LETTER SH
0F PAHLAVI LETTER GH

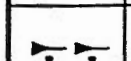
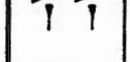
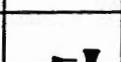
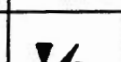
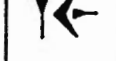
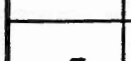
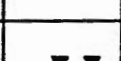
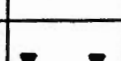
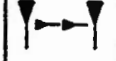
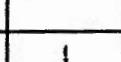
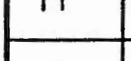
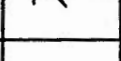
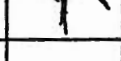
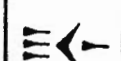
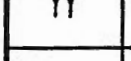
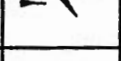
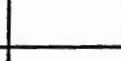
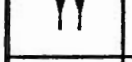
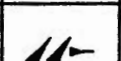
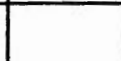
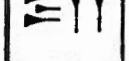
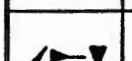
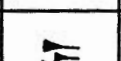
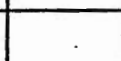
10 PAHLAVI LETTER F
11 PAHLAVI LETTER K
12 PAHLAVI LETTER G
13 PAHLAVI LETTER L
14 PAHLAVI LETTER Y
15 PAHLAVI LETTER M
16 PAHLAVI LETTER N
17 PAHLAVI LETTER N OVERDOT
18 PAHLAVI LETTER N ACUTE
19 PAHLAVI LETTER N TILDE
1A PAHLAVI LETTER V
1B PAHLAVI LETTER H
1C PAHLAVI LETTER H OGONEK
1D PAHLAVI LETTER E
1E PAHLAVI LETTER O
1F PAHLAVI LETTER HW

20 PAHLAVI LETTER AA
21 PAHLAVI LETTER I
22 PAHLAVI LETTER II
23 PAHLAVI LETTER U
24 PAHLAVI LETTER UU
25 PAHLAVI LETTER SCHWA
26 PAHLAVI LETTER SCHWA SCHWA
27 PAHLAVI LETTER EE
28 PAHLAVI LETTER OO
29 PAHLAVI LETTER A OGONEK
2A PAHLAVI LETTER W

2B PAHLAVI LETTER SH
2C PAHLAVI LETTER ZH
2D PAHLAVI FULL STOP

Old Persian Cuneiform

draft
DEC 10, 1991

0								
1								
2								
3								
4								
5								
6								
7								
8								
9								
A								
B								
C								
D								
E								
F								

Old Persian Cuneiform

Old Persian cuneiform was used extensively over a large area drained by the Euphrates and Tigris rivers in lands that were once called Akkad and Sumer. It was the first type of cuneiform to be deciphered in modern times. The script is traditionally said to have been invented by Darius I (ca 521-486 BC) so that he might be comparable to Babylonian and Assyrian kings; by about 300 BC it had fallen out of use.

Old Persian inscriptions were first seriously studied by C. Niebuhr in 1765, though various types of cuneiform inscriptions had been known in the West for quite some time. Preliminary studies which eventually culminated in decipherment and understanding of the language were made as early as 1798 by O.G. Tycheson and F.C.C. Münter; they were succeeded in the task by G.F. Grotefend and others. Decipherment was essentially complete by about 1845. Decipherment was also achieved, quite independently, by H. C. Rawlinson between about 1836 and 1850. A rather small literature in Old Persian is extant, but it includes some lengthy carved inscriptions at Behistun and Persepolis (northeast of modern Baghdad along the Tigris River).

The system is essentially a syllabary of thirty-six signs, augmented by a specialized word divider and five ideographs. The ideographs are for king, country, earth, god, and the supreme diety of the time, Ahura-Mazda. Of these, the latter appears in several minor glyphic variations. The script is thought to be complete in this encoding; it should not be confused with the much earlier ideographic cuneiform scripts of Akkadian and Sumerian derivation.

Issues: The numbers (1, 2, 3, 10, 20, 40, 100) may be incomplete in the chart, but sufficient information is not available at this time. These numbers could be compressed together, but in this chart are spread out into what may be appropriate places, assuming the existence of other number signs. They could also be packed at the end of the script. If a word-divider is shared with Ugaritic Cuneiform (and was encoded there), then the seven numbers could be put into the third column of the chart, and Old Persian would fit into three complete rows instead of taking part of a fourth row.

Some Sources

Cleator, P. E. *Lost Languages.*
Friedrich, Johannes. *Extinct Languages.*
Coulmas, Florian. *Writing Systems of the World.*

Rev 92/10/20

Old Persian Names List, draft Dec 10, 1991

00 OLD PERSIAN CUNEIFORM LETTER A
01 OLD PERSIAN CUNEIFORM LETTER I
02 OLD PERSIAN CUNEIFORM LETTER U
03 OLD PERSIAN CUNEIFORM LETTER BA
04 OLD PERSIAN CUNEIFORM LETTER CA
05 OLD PERSIAN CUNEIFORM LETTER CHA
06 OLD PERSIAN CUNEIFORM LETTER DA
07 OLD PERSIAN CUNEIFORM LETTER DI
08 OLD PERSIAN CUNEIFORM LETTER DU
09 OLD PERSIAN CUNEIFORM LETTER FA
0A OLD PERSIAN CUNEIFORM LETTER GA
0B OLD PERSIAN CUNEIFORM LETTER GU
0C OLD PERSIAN CUNEIFORM LETTER HA
0D OLD PERSIAN CUNEIFORM LETTER HHA
0E OLD PERSIAN CUNEIFORM LETTER JA
0F OLD PERSIAN CUNEIFORM LETTER JI

10 OLD PERSIAN CUNEIFORM LETTER KA
11 OLD PERSIAN CUNEIFORM LETTER KU
12 OLD PERSIAN CUNEIFORM LETTER LA
13 OLD PERSIAN CUNEIFORM LETTER MA
14 OLD PERSIAN CUNEIFORM LETTER MI
15 OLD PERSIAN CUNEIFORM LETTER MU
16 OLD PERSIAN CUNEIFORM LETTER NA
17 OLD PERSIAN CUNEIFORM LETTER NU
18 OLD PERSIAN CUNEIFORM LETTER PA
19 OLD PERSIAN CUNEIFORM LETTER RA
1A OLD PERSIAN CUNEIFORM LETTER RU
1B OLD PERSIAN CUNEIFORM LETTER SA
1C OLD PERSIAN CUNEIFORM LETTER SHA
1D OLD PERSIAN CUNEIFORM LETTER TA
1E OLD PERSIAN CUNEIFORM LETTER TU
1F OLD PERSIAN CUNEIFORM LETTER THA

20 OLD PERSIAN CUNEIFORM LETTER WA
21 OLD PERSIAN CUNEIFORM LETTER WI
22 OLD PERSIAN CUNEIFORM LETTER YA
23 OLD PERSIAN CUNEIFORM LETTER ZA
24 OLD PERSIAN CUNEIFORM WORD DIVIDER
25 OLD PERSIAN CUNEIFORM IDEOGRAPH KING
26 OLD PERSIAN CUNEIFORM IDEOGRAPH COUNTRY
27 OLD PERSIAN CUNEIFORM IDEOGRAPH EARTH
29 OLD PERSIAN CUNEIFORM IDEOGRAPH GOD
2A OLD PERSIAN CUNEIFORM IDEOGRAPH AHURA-MAZDA

Early Phoenician

araiL
DEC 10, 1991

0	𐤀	𐤁						
1	𐤂	𐤃						
2	𐤄	𐤅						
3	𐤆	𐤇						
4	𐤈	𐤉						
5	𐤊	𐤋						
6	𐤌							
7	𐤍							
8	𐤎							
9	𐤏							
A	𐤐							
B	𐤑							
C	𐤒							
D	𐤓							
E	𐤔							
F	𐤕							

Phoenician

The Phoenician alphabet and its successors were widely used over a broad area surrounding the Medierranean Sea. Phoenician evolved over several hundred years from the end of the 2nd millenium BC (before 1100 BC) with some modifications until the 2nd century BC, with the last neo-Punic inscriptions dating from about the 3rd century AD. The Phoenician alphabet is a forerunner of the Etruscan, Latin, Greek, Arabic, Hebrew, and Syriac scripts among others, many of which are still in modern use. It has also been suggested that Phoenician is the ultimate source of the Indic scripts descending from Brahmi and Kharoshthi.

Phoenician is quintessentially illustrative of the historical problem of where to draw lines in an evolutionary tree of contiuously changing scripts extending over thousands of years. The twenty two letters in the Phoenician block may be used, with appropriate font changes, to express Early Phoenician, Moabite, Early Hebrew, Later Phoenician, and Punic, and possibly some Early Aramaic. It is especially intended for use with Phoenician and Punic. The historical cut that has been made in Unicode considers the line from Phoenician to Punic to represent a single continuous branch of script evolution.

Phoenician is generally written from right to left horizontally. Phoenician language inscriptions usually have no space between words; there are sometimes dots between words in later inscriptions (e.g., in Moabite inscriptions). Typical fonts for the Phoenician and especially Punic have very exaggerated descenders. These descenders help distinguish the main line of Phoenician evolution toward Punic from the other (e.g., Hebrew) branches of the script, where the descenders instead grew shorter over time.

Some Sources

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

Rev 92/10/30

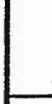
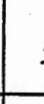
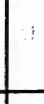
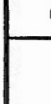
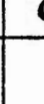
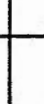
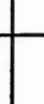
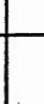
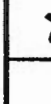
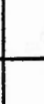
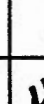
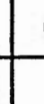
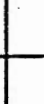
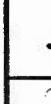
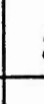
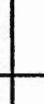
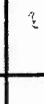
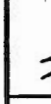
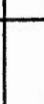
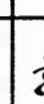
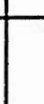
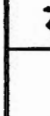
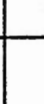
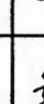
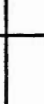
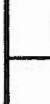
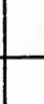
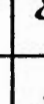
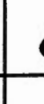
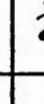
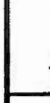
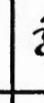
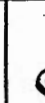
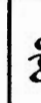
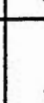
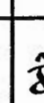
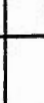
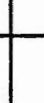
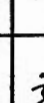
Early Phoenician Names List, draft Dec 10, 1991

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

10 EARLY PHOENICIAN LETTER SAN
11 EARLY PHOENICIAN LETTER QOPPA
12 EARLY PHOENICIAN LETTER RESH
13 EARLY PHOENICIAN LETTER SHIN
14 EARLY PHOENICIAN LETTER TAU
15 EARLY PHOENICIAN LETTER WAW

RÓNG (LEPCHA)

DRAFT

	0	1	2	3	4			
0								
1								
2								
3								
4								
5								
6								
7								
8								
9								
A								
B								
C								
D								
E								
F								

Róng (Lepcha)

The Róng script (also called *Lepcha*) is used to write the Róng language of Sikkim (located between Nepal and Bhutan, just south of Tibet). It bears structural similarity to Tibetan, from whence it probably ultimately derives. The script is tradtionally held to have been invented by a Sikkim Raja (named *Phyag-rdor-rnam-rgyal*) in the early 18th century. This "invention" was probably actually an extensive revision of an older script. A unique feature of the script is its use of syllable-final "floating consonant signs" (U+xx37 → U+xx3F). These signs were probably invented for and introduced into the Róng script by the reviser. This structural feature eliminates the need for any conjunct consonants in Róng. The signs for letters with an infixed "L" sound are likewise unknown from other scripts of the area, and seem to be a unique feature.

The two signs *KYA* and *KRA* (U+xx24 and U+xx25) are analogous to the Tibetan *ya-ta* and *ra-ta* but are affixed after the preceding consonant rather than as subscripts. Róng typography uses a number of very regular ligatures formed by consonants with succeeding *KYA* and *KRA*. There is also a special ligature form of *KRA* followed by *KYA*, which itself forms ligatures with the preceding consonant. Of the seven vowel signs, three (U+xx31 → U+xx33) are reordered in display, as are two of the syllable-final floating consonant signs (U+xx3E and U+xx3F). When a vowel sign of the reordering type is followed by one of the floating consonant signs of the reordering type, the consonant sign is written to the *left* of the vowel sign.

Róng occasionally makes use of a floating dot (U+xx2E) below consonants to distinguish special pronunciations (an innovation introduced by Mainwaring). The floating mark *RAN* (U+xx2F) is used over consonants (and above their associated floating consonant signs, if any) to indicate a slight lengthening or emphasis of the vowel. The only punctuation is U+xx2D, equivalent to the Devanagari *danda*. Róng seems to always be written with space between words or compound words.

Issues: Unless there has been a recent revival, this script is probably not in active use at all as of this writing (1992). Haarh's 1959 article seems to imply that the script was still in use at that time. The Baptist mission in the late 1800s apparently printed three books of the New Testament in the script. While Mainwaring's work (1876) gives an encouraging picture, Gorer's ethnography of the Lepcha (written in 1938, revised in 1967) is quite clear as regards the script. Gorer contends that it

was rather artificially revived by the eccentric General Mainwaring, and reports that he could find only one old lama who possessed or could read a book in the script:

> ...the Lepcha script, never widely known, has now completely fallen into disuse; in order to read the scriptures Lepchas have to learn a new, and otherwise completely useless, alphabet; most of them are far more familiar with Nepali. ... All the existing Lepcha manuscripts of which I have heard are translations of the Tibetan lamaist scriptures... (Gorer, p. 38-39)

Róng is structurally similar to Kirat (Limbu), especially in its use of floating final consonant signs, which are also used in Kirat . In this respect the two scripts *differ* from most (or all?) other scripts of the area. These signs would seem to be an innovation of the Róng script which was taken up in the Kirat script. The language for which the script was originally invented is a "mono-syllabic" type language. The script is apparently derived from the Tibetan script, but Róng was revised in the early 1700s, at which time these signs were introduced. This model presumes the final consonant signs to be a unique invention that makes structural sense in the script and the language which it is intended to serve. In this author's view, this model is straightforward, and should be more or less retained unless strong evidence to the contrary becomes available.

It has been argued elsewhere, however, that the Róng (and Kirat) final consonants are simply rendering forms, and hence should be spelled by means of an affixed invisible *virama* (which would follow a normal consonant and produce visually one of the floating signs in word-final position). No evidence available at this time suggests that any type of *virama* (visible or invisible) is known in the script at all. The possibility cannot be completely discounted, however, since the script derives ultimately from Brahmi and the other Indic scripts, and there is some evidence for an invisible *virama* (at least conceptually) in Tibetan. Such a model would include a *virama* and use it to spell the final consonant signs; it would also presumably encode the consonants with infix-l offglide (such as HLA) with this virama as well. Such a model is not without some merit, chiefly in paralleling existing script encodings.

It has also been suggested that Róng (as well as Kirat) could be encoded (at least partially) parallel to the order of the Tibetan block, or it could be encoded parallel to ISCII. While neither of these is particularly compelling, the closer relation to the Tibetan script makes it the more likely choice, if it must be encoded parallel to another script.

The letters with infixed "L" could also be moved elsewhere in the alphabetic order, which may make alphabetization easier or more clear. Mainwaring's dictionary order may be artificial.

This draft for Róng is by no means a final answer. The available sources are somewhat sketchy as regards fine points of the script; not enough analytical sources or textual sources are available at this time to conclusively resolve some of the issues. See also the block introduction for Kirat (Limbu).

Some Sources

Mainwaring, G. B. *A Grammar of the Róng (Lepcha) Language.*
Mainwaring, G. B. *Dictionary of the Lepcha Language.*
Haarh, Erik. *The Lepcha Script.*
Gorer, Geoffrey. *Himalayan Village.*

Rev 92/11/25

Draft RONG/LEPCHA Names List, rev 10/21/92.

00 RONG/LEPCHA LETTER KA
01 RONG/LEPCHA LETTER KHA
02 RONG/LEPCHA LETTER GA
03 RONG/LEPCHA LETTER NGA
04 RONG/LEPCHA LETTER CHA
05 RONG/LEPCHA LETTER CHHA
06 RONG/LEPCHA LETTER JA
07 RONG/LEPCHA LETTER NYA
08 RONG/LEPCHA LETTER TA
09 RONG/LEPCHA LETTER THA
0A RONG/LEPCHA LETTER DA
0B RONG/LEPCHA LETTER NA
0C RONG/LEPCHA LETTER PA
0D RONG/LEPCHA LETTER PHA
0E RONG/LEPCHA LETTER FA
0F RONG/LEPCHA LETTER BA

10 RONG/LEPCHA LETTER MA
11 RONG/LEPCHA LETTER TSA
12 RONG/LEPCHA LETTER TSHA
13 RONG/LEPCHA LETTER ZA
14 RONG/LEPCHA LETTER YA
15 RONG/LEPCHA LETTER RA
16 RONG/LEPCHA LETTER LA
17 RONG/LEPCHA LETTER HA
18 RONG/LEPCHA LETTER VA
19 RONG/LEPCHA LETTER SA
1A RONG/LEPCHA LETTER SHA
1B RONG/LEPCHA LETTER WA
1C RONG/LEPCHA LETTER KLA
1D RONG/LEPCHA LETTER GLA
1E RONG/LEPCHA LETTER PLA
1F RONG/LEPCHA LETTER FLA

20 RONG/LEPCHA LETTER BLA
21 RONG/LEPCHA LETTER MLA
22 RONG/LEPCHA LETTER HLA
23 RONG/LEPCHA LETTER A
24 RONG/LEPCHA Affix KYA
25 RONG/LEPCHA Affix KRA
26 unencoded
27 unencoded
28 unencoded
29 unencoded
2A unencoded

2B unencoded
2C unencoded
2D RONG/LEPCHA FINAL PUNCTUATION (DANDA)
2E RONG/LEPCHA DOT BELOW
2F RONG/LEPCHA NON-SPACING SIGN RAN

30 RONG/LEPCHA VOWEL SIGN AA
31 RONG/LEPCHA VOWEL SIGN I
32 RONG/LEPCHA VOWEL SIGN O
33 RONG/LEPCHA VOWEL SIGN OO
34 RONG/LEPCHA VOWEL SIGN U
35 RONG/LEPCHA VOWEL SIGN UU
36 RONG/LEPCHA VOWEL SIGN E
37 RONG/LEPCHA FINAL CONSONANT SIGN AK
38 RONG/LEPCHA FINAL CONSONANT SIGN AM
39 RONG/LEPCHA FINAL CONSONANT SIGN AL
3A RONG/LEPCHA FINAL CONSONANT SIGN AN
3B RONG/LEPCHA FINAL CONSONANT SIGN AB
3C RONG/LEPCHA FINAL CONSONANT SIGN AR
3D RONG/LEPCHA FINAL CONSONANT SIGN AT
3E RONG/LEPCHA FINAL CONSONANT SIGN NG
3F RONG/LEPCHA FINAL CONSONANT SIGN ANG

40 RONG/LEPCHA DIGIT ZERO
41 RONG/LEPCHA DIGIT ONE
42 RONG/LEPCHA DIGIT TWO
43 RONG/LEPCHA DIGIT THREE
44 RONG/LEPCHA DIGIT FOUR
45 RONG/LEPCHA DIGIT FIVE
46 RONG/LEPCHA DIGIT SIX
47 RONG/LEPCHA DIGIT SEVEN
48 RONG/LEPCHA DIGIT EIGHT
49 RONG/LEPCHA DIGIT NINE

Composite Chart of Northern Runes (Rev 92/10/22)

Notes on the Runic Chart

This proposed composite block is based on a preliminary analysis of elements that clearly need to be *distinguished* within any one of the four idealized Runic alphabets (shown below). Some outstanding distinctions are these:

Runes 4a, 5, 7a both occur in the Anglo-Saxon
Runes 4b, 6a both occur in the Danish
Runes 10c occurs as a variant of 20a in the Anglo-Saxon
Rune 19a is "m" in the Danish, "R" (?) in the Germanic, "x" in the Anglo-Saxon
Runes 13, 14a both occur in the Anglo-Saxon
Runes 21b, 25 both occur in Swedo-Norwegian (whereas elsewhere they might be used interchangeably for "l" in retrograde inscriptions)

Northern Runes

The Northern Runic script was widely used in northern Europe, primarily in Scandinavia and Germany, between about the second and eleventh centuries AD when it was gradually replaced by the Latin alphabet. (We call it the Northern Runic script to distinguish it from other so-called Runic scripts, such as the Turkic.) Northern Runes were also used in England from about the 7th century AD. Some 5000 known Runic inscriptions survive from the central cultural area and outlying areas as far away as Russia, Poland, and North America. Inscriptions are found primarily on wood, stone, and metal objects, but there are also extant manuscripts that explain the runes. These inscriptions often consist simply of the letters of the (local) alphabet written out in standardized order, so the alphabetical orders are well known and various stages can be compared with relative ease.

The Runic alphabet for a given language and locale is commonly referred to as the *futhark*, a name derived from its first six letters. There are two major branches of Northern Runes, the Germanic branch and the Scandinavian branch, which differ in their arrangement and in the forms of many characters. The Runic script modelled in this block is a minimal *composite* of graphic forms derived from the major Runic alphabets. These alphabets and their glyphic variants are considered here to be built from elements of a single larger Runic script. The Runic script, however, is not a predefined entity, rather a theoretical construction consisting of the graphic elements which must be minimally distinguished and grouped into "glyphic alternative" bundles where appropriate.

The Scandinavian *futhark* consisted of 16 base characters, apparently derived by eliminating symbols from the older *futhark*, but with other changes as well. A dot or double-dot mark was used on five of these base characters bringing the total distinct symbols to 21. In several instances the *form* used for one sound in the Scandinavian was used for a *different* sound in the Germanic (this fact is more apparent when various *futharks* using variant glyphs are brought together for comparison than it is in the charts shown here). The Scandinavian *futhark* includes the so-called "short twig" or Hälsing Runic shapes.

The Runes evolved considerably over the course of some 1000 years, often differently in various locales. It cannot be stressed enough that the Unicode Runic block is *abstracted* from the historical inscriptions used throughout the Runic

cultural area. Some characters, our composite runes numbered 10 and 26 for instance, assumed a wide variety of related forms; the *h* rune (composite number 13) could have one or two bars. The glyphic forms used in the charts are *not* intended to be normative, merely illustrative of the more typical shapes.

Display and rendering: The predominant writing direction was in horizontal lines from left to right. However, they were also sometimes written retrograde. The earliest inscriptions were written with no punctuation and run-together words, much like ancient Greek. Later inscriptions often made use of a colon (:) or middle-dot between words (not included in this block). Fonts for the Runes would probably encode a superset of the most widely used glyphs, from which glyphs would be chosen to represent one or the other of the desired *futhark* surface structures with their variations. (The stroke font designed for the accompanying chart is one example; the full glyphic complement of this font is shown.) Some later inscriptions also mixed Latin letters with runes, so it seems not unreasonable that the most flexible fonts would include various harmonious Latin shapes as well. Ligatures were sometimes used in Runic inscriptions. They seem to have been freely formed by bodily fusion of two or more characters, especially fusion of the vertical stroke components. No such ligatures are encoded in this block.

Issues: Because the Anglo-Saxon and Germanic *futharks* are closely related in most of their forms and functions, the major part of the Anglo-Saxon one can be mapped directly onto the Germanic *futhark* of 24 letters. (There are seven extra characters used for Anglo-Saxon.) The Runic block could then be divided into two parts, one representing the Anglo-Saxon and Germanic branches with a total of 31 characters (referred to as the *older futhark*), and another representing the Scandinavian branch of fewer characters with some different forms (referred to as the *younger futhark*). Division in this manner (encoding two separate sections of 31 and 24 characters) can be easily envisioned by comparing the four alphabets shown in the accompanying chart. Another obvious alternative would be to encode the entire set on phonemic principles (with minor variations), which would be equivalent (or nearly so) to a simple interwoven unification of the four aforementioned alphabets. All of the approaches seem to have disadvantages.

We here use the comparative Runic sets on the following pages (after Healey). One inconsistency introduced by division into two blocks is that the 4th Germanic rune (our composite number 4a) must still be distinguished from the 4th Anglo-Saxon rune (our number 7). Anglo-Saxon puts the Germanic 4th rune shape

at its 26th location). The only choice is to put one or the other out of alphabetical order. There are several other minor problems with the division, notably that our rune (composite number) 19a is used for two or more different sounds.

Implementation of Runes almost requires some standard method of indicating glyphic preference, as many of the Runic shapes seem to be free variants that probably make a great deal of difference to scholars, though legibility should not be impaired if normative forms are used.

Some Sources

Page, R. I. *Runes.*
Antonsen, Elmer H. *The Runes: The Earliest Germanic Writing System.*
Xerox Character Code Standard.
Haugen, Einar. *History of the Scandinavian Languages.*
??? pages from "runläsboken" (in Swedish).

Rev 92/11/25

Full Glyph Complement (of this particular font, excluding the 'space' glyph):

ᚠ ᚢ ᚦ ᚨ ᚩ ᚭ ᛆ ᚫ ᚬ ᚱ ᛦ ᚳ ᚲ ᚷ ᚴ ᚹ ᚻ ᚺ ᚽ ᚾ ᛆ ᛁ ᛄ ᛃ ᛇ ᛇ ᛈ ᛔ ᛉ ᛋ ᛊ ᛂ
ᚴ ᛏ ᛐ ᛒ ᛓ ᛖ ᛗ ᛧ ᛚ ᛝ ᛜ ᛙ ᛟ ᛞ ᛏ ᛞ ᚤ ᚪ ᛠ ᛡ ᛤ ı ·

Germanic:

ᚠ ᚢ ᚦ ᚨ ᚱ ᚲ ᚷ ᚹ ᚺ ᚾ ᛁ ᛃ ᛇ ᛈ ᛉ ᛊ ᛏ ᛒ ᛖ ᛗ ᛚ ᛜ ᛟ ᛞ

Anglo-Saxon:

ᚠ ᚢ ᚦ ᚩ ᚱ ᚳ ᚷ ᚹ ᚻ ᚾ ᛁ ᛄ ᛇ ᛈ ᛉ ᛋ ᛏ ᛒ ᛖ ᛗ ᛚ ᛝ ᛟ ᛞ ᚪ ᚫ ᚣ ᛠ ᛡ ᛣ ᛤ

Danish:

ᚠ ᚢ ᚦ ᚬ ᚱ ᚴ ᚼ ᚾ ᛁ ᛅ ᛋ ᛏ ᛒ ᛘ ᛚ ᛦ (ᛂ ᚵ ᚤ ᛑ ᛔ)

Swedo-Norwegian:

ᚠ ᚢ ᚦ ᚭ ᚱ ᚴ ᚽ ᚿ ᛁ ᛆ ᛌ ᛐ ᛓ ᛙ ᛚ ᛧ (ᛂ ᚵ ᚤ ᛑ ᛔ)

South Arabian

draft
DEC 10, 1991

	0	1						
0								
1								
2								
3								
4								
5								
6								
7								
8								
9								
A								
B								
C								
D								
E								
F								

Epigraphic South Arabian

The script known as South Arabian is related to the Proto-Canaanite and early Semitic alphabets, but the shapes are remarkably unique for such a derivation. It is also an ancestor of the modern Ethiopic script. Inscriptions in this script are found in Southern Arabia (ancient Sabaean and Minaean kingdoms) dating from as far back as 500 BC. The script was apparently used until about 600 AD. According to Healey (see below), the alphabetic order has been reconstructed on fragmentary evidence. The order given here follows that given by Healey.

The letters as 10 and 11 probably correspond to the Arabic *hamzah* and *ain*, but this is not certain from information currently available.

Issues: The South Arabian alphabet could be arranged parallel to the Semitic alphabets. See the introduction to the Early Alphabet blocks for further discussion.

Some Sources

Healey, John F. *The Early Alphabet.*

Rev 92/10/29

Epigraphic South Arabian, draft names 92/10/20

00 SOUTH ARABIAN LETTER H
01 SOUTH ARABIAN LETTER L
02 SOUTH ARABIAN LETTER H UNDERDOT
03 SOUTH ARABIAN LETTER M
04 SOUTH ARABIAN LETTER Q
05 SOUTH ARABIAN LETTER W
06 SOUTH ARABIAN LETTER S HACEK
07 SOUTH ARABIAN LETTER R
08 SOUTH ARABIAN LETTER B
09 SOUTH ARABIAN LETTER T
0A SOUTH ARABIAN LETTER S
0B SOUTH ARABIAN LETTER K
0C SOUTH ARABIAN LETTER N
0D SOUTH ARABIAN LETTER H UNDERBAR
0E SOUTH ARABIAN LETTER S ACUTE
0F SOUTH ARABIAN LETTER F

10 SOUTH ARABIAN LETTER RIGHT HALF RING (HAMZAH)
11 SOUTH ARABIAN LETTER LEFT HALF RING (AIN)
12 SOUTH ARABIAN LETTER D UNDERDOT
13 SOUTH ARABIAN LETTER G
14 SOUTH ARABIAN LETTER D
15 SOUTH ARABIAN LETTER G ACUTE
16 SOUTH ARABIAN LETTER T UNDERDOT
17 SOUTH ARABIAN LETTER Z
18 SOUTH ARABIAN LETTER D UNDERBAR
19 SOUTH ARABIAN LETTER Y
1A SOUTH ARABIAN LETTER T UNDERBAR
1B SOUTH ARABIAN LETTER S UNDERDOT
1C SOUTH ARABIAN LETTER Z UNDERDOT

SYRIAC (ESTRANGELO) 92/10/29

							Mandaic	
0	ܐ	ܓ					ࡀ	ࡂ
1	ܒ	ܕ					ࡁ	ࡃ
2	ܓ	ܗ					ࡂ	ࡄ
3	ܕ	ܘ					ࡃ	ࡅ
4	ܗ	ܙ					ࡄ	ࡆ
5	ܘ	ܚ					ࡅ	ࡇ
6	ܙ						ࡆ	ࡈ
7	ܚ						ࡇ	
8	ܛ						ࡈ	
9	ܝ						ࡉ	
A	ܟ/ܟ						ࡊ	
B	ܠ						ࡋ	
C	ܡ						ࡌ	
D	ܢ						ࡍ	
E	ܣ						ࡎ	
F	ܥ						ࡏ	

Syriac

The Syriac script is a later descendent of the Aramaic script. The earliest known Syriac inscriptions are dated about 6 AD from near the town of Edessa to write the Aramaic dialect that became Syriad. The Syriac script really represents a family of three closely related writing styles called *Estrangela*, *Nestorian*, and *Serta* (the latter is also called *Jacobite*). The earliest form that became distinguished from Aramaic itself is *Estrangela*, developed about the 5th century AD. It was used extensively from the earliest times to record various Christian scriptures. The Syriac script is still in modern use. According to Healey (1990):

> "Syriac speaking communities have survived in large numbers in the area around the point where the borders of Syria, Turkey, and Iraq meet, and there are also *emigré* communities in Europe and the United States. Books, magazines and newspapers are still produced in the Syriac scripts."

The Syriac scripts are generally cursive or semi-cursive, with some letters joining regularly to others and sometimes changing shape in a manner similar to the Arabic script. Vowel signs are known to exist, but available sources do not discuss them.

Issues: The vowel signs at least must be added to complete the Syriac proposal. There seem to be at least two different non-spacing vowel systems: one is attributed to Jacob of Edessa and utilizes small letters written above or below others to indicate following vowels; the other is an older dotting system.

The chart shows in parallel the Mandaic alphabet (which includes the extra letter *e* at the end). It is not clear whether Mandaic should be unified with the Syriac block or not; it might be better encoded using the Aramaic block, or encoded separately.

Note that this order differs from the Early Phoenician and Aramic orders. It is not known whether *waw* in particular should come at the end, or at its place here.

Some Sources

Healey, John F. *The Early Alphabet.*
Diringer, David. *Writing.*

Syriac Names List, draft 92/10/29

00 SYRIAC LETTER ALAP
01 SYRIAC LETTER BET
02 SYRIAC LETTER GAMAL
03 SYRIAC LETTER DALAT
04 SYRIAC LETTER HE
05 SYRIAC LETTER WAW
06 SYRIAC LETTER ZAYN
07 SYRIAC LETTER HET
08 SYRIAC LETTER TET
09 SYRIAC LETTER YO
0A SYRIAC LETTER KAP
0B SYRIAC LETTER LAMAD
0C SYRIAC LETTER MIM
0D SYRIAC LETTER NUN
0E SYRIAC LETTER SEMKAT
0F SYRIAC LETTER E

10 SYRIAC LETTER PE
11 SYRIAC LETTER SADE
12 SYRIAC LETTER QOP
13 SYRIAC LETTER RES
14 SYRIAC LETTER SIN
15 SYRIAC LETTER TAW

TAGALOG 92/10/28

	Tagalog			Tagbanuwa			Mangyan	
0								
1								
2								
3								
4								
5								
6								
7								
8								
9								
A								
B								
C								
D								
E								
F								

Tagalog and Mangyan (Buhid)

Tagalog is a script of the Philippines. It was formerly used to write the Tagalog, Bisaya, Iloko, and other languages. The Tagalog language is very much alive, but now utilizes the Latin script. The Tagalog script is distantly related to the scripts of the southern Indian subcontinent, but the exact route by which they were brought to the Philippines is not certain. It seems that they may have been transported by way of the palaeographic scripts of Western Java between the 10th and 14th centuries. Written accounts of the Tagalog script by Spanish missionaries, and documents in Tagalog, are known from about the period of initial Spanish incursion (mid-1500s). It has (or had) two living descendents – the *Mangyan* and *Tagbanuwa* scripts – both of which will be covered below.

Vowel signs are used in a manner similar to that employed by the scripts of the Indian subcontinent, from whence Tagalog seems to derive. The vowel I is written with a mark above, and the vowel U with an identical mark below the associated consonant. The mark looks like the sign "**>**". It is known as *kulit* or *tulbok* in Mangyan and *ulitan* in Tagbanuwa. The script has *only* the two vowel signs I and U, which are also used respectively to stand for the vowels E and O. Though all languages normally written with this script have syllables possessing final consonants, they cannot be expressed in the script. Reforms to express final consonants or to add the missing vowel signs were apparently proposed at various times, but were always rejected by native users who considered the script adequate. Native speakers of Tagbanuwa, for instance, apparently have no trouble distinguishing uses of the vowel sign I for the vowel *e*, or the sign U for *o*. In Tagalog there are several similar glyphs for the independent vowel I; it could variously be written with the wavy line above or below the straight one.

Tagalog is *read* from left to right in horizontal lines running from top to bottom. It may be *written* either in that manner, or in vertical lines running from bottom to top, moving from left to right. In the latter case, the letters are written sideways so they may be read horizontally. This method of writing may be due to the medium and writing implements used. It was often scratched with a sharp instrument onto beaten strips of bamboo which were held pointing away from the body and worked from the proximal to distal ends, from left to right.

Between words in Tagalog, a sign similar to double *danda* seems to be used (see the example in Nakanishi). The double *danda* is not included in the chart.

The alphabetical order of Tagalog is known from Tagbanuwa speakers and is described in folktales. This order is used in the accompanying charts. The two vowel signs are added at the end of the alphabet.

The accompanying chart is divided into three segments. The leftmost group are the forms used for classical Tagalog. The middle group, exactly paralleling the Tagalog, are the forms used for Tagbanuwa. The rightmost group are the forms used for Mangyan.

Tagbanuwa: The Tagbanuwa letter forms are nearly the same as the old Tagalog forms, and the lineage is obvious as can be seen from the accompanying charts. Particularly different are the letters I and KA. Modern Tagbanuwa does not use the letter HA, hence this spot is left blank in the Tagbanuwa chart.

Mangyan: Mangyan is the term given to the Bongabon Mangyans, also known as Buhid or Bukid. The Mangyan letter forms differ significantly from their Tagalog counterparts. They were normally incised on bamboo, and the influence of the medium is unmistakably expressed in the angular letter forms. The vowel signs I and U are normally written as strokes *attached to* the main body of the associated consonant, in contrast to the Tagalog case for the same vowel signs. A font for Mangyan might thus be completely "unrolled" as a syllabary, requiring about 50 distinct glyphs.

Issues: It is known that Tagbanuwa and Mangyan were being actively used as recently as the early 1960s, as near as can be ascertained from evidence in Francisco's monograph. It is not known whether they are still being used as of this date (1992). It is unclear whether to classify them (and thus Tagalog) as living or extinct scripts. The extent to which their encoding is important to living communities is likewise uncertain.

Mangyan should perhaps be separately encoded from a Tagalog & Tagbanuwa block due to (1) significant differences in nearly all letter forms and (2) the means by which vowel signs are attached and (3) as the two scripts are (or were) living side by side there may be a need for distinguishing them in plaintext, (4) either one may not be readable by those unfamiliar with the other.

Some Sources

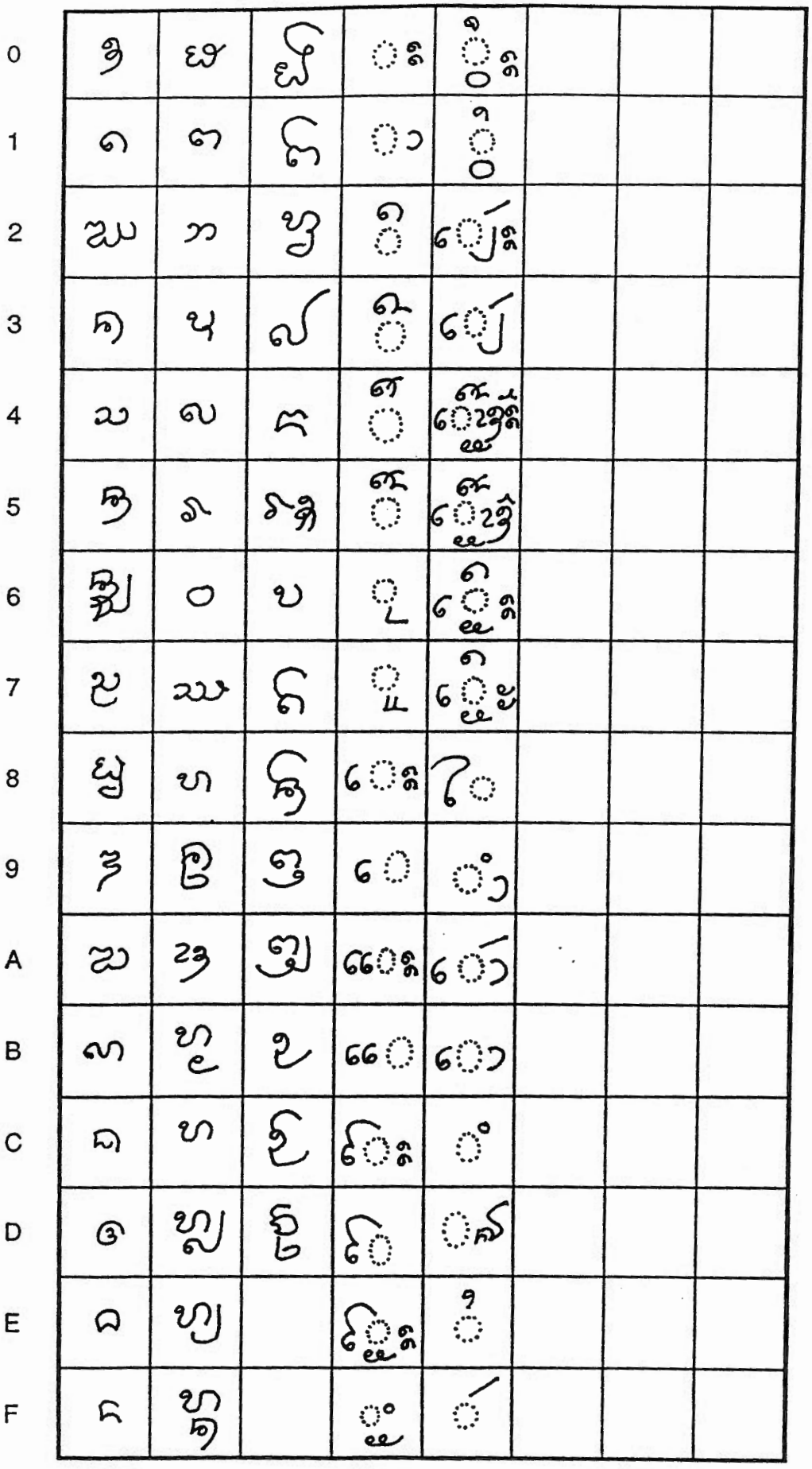
Francisco, Juan R. *Philippine Palaeography.*
Faulmann, Carl. *Schriftzeichen und Alphabete aller Zeiten und Völker.*

Rev 92/10/29

Tagalog Names, draft 92/10/21

00 TAGALOG LETTER A
01 TAGALOG LETTER I AND E
02 TAGALOG LETTER U AND O
03 TAGALOG LETTER BA
04 TAGALOG LETTER DA
05 TAGALOG LETTER GA
06 TAGALOG LETTER HA
07 TAGALOG LETTER KA
08 TAGALOG LETTER LA
09 TAGALOG LETTER MA
0A TAGALOG LETTER NA
0B TAGALOG LETTER NGA
0C TAGALOG LETTER PA
0D TAGALOG LETTER SA
0E TAGALOG LETTER TA
0F TAGALOG LETTER WA
10 TAGALOG LETTER YA
11 TAGALOG VOWEL SIGN I
12 TAGALOG VOWEL SIGN U

Similarly for Mangyan, if separately encoded:
XX MANGYAN LETTER XX



Tai Lu (Chieng Mai, Northern Thai)

The Tai Lu script is widely used for various Tai dialects in northern Thailand, Yunnan, and parts of Burma (they are variously referred to as Lannathai, Yuan, or Kam Muang). The Tai Lu script is of the Indic variety, and is structurally similar to both the Thai and Burmese 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 Lu script, and hence included here.

The *language* known as Tai Lu is in use in northern Thailand and in Yunnan province of China. There are about 1 million living speakers of Tai Lu, and this script is officially recognized by the Chinese government.

Each Tai Lu 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.

Some Sources

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

Rev 92/11/25

Tai Lu (Chieng Mai, Northern Thai) names, rev 92/10/21

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

TAI MAU 92/10/29

0	ᥐ	ᥑ	ᥒ					
1	ᥓ	ᥔ	ᥕ					
2	ᥖ	ᥗ	ᥘ					
3	ᥙ	ᥚ	ᥛ					
4	ᥜ	ᥝ	ᥞ					
5	ᥟ	ᥠ	ᥡ					
6	ᥢ	ᥣ	ᥤ					
7	ᥥ	ᥦ	ᥧ					
8	ᥨ	ᥩ						
9	ᥪ	ᥫ						
A	ᥬ	ᥭ						
B	᥮	᥯						
C	ᥰ	ᥱ						
D	ᥲ	ᥳ						
E	ᥴ	᥵						
F	᥶	᥷						

Tai Mau, Tai Nua

The Tai Mau or Tai Nua script is a recent invention that is reported to have been in use only since 1940. It is apparently used for writing several Shan languages within China (Yunnan) and Northeastern Burma (between the Nam Mau and Salween rivers). The Tai Mau script was invented (revised?), apparently, as a reaction to a reported revision of another script used by the Tai Tai (Burma).

This script is remarkably simpler in structure than those used for standard Thai and Northern Thai (see Thai and Tai Lu block introductions). It has many different attributes when considered as a relative of those scripts, mostly in the features which it lacks: it has no non-spacing tone marks, non-spacing vowel signs, re-ordering matras, or conjunct consonant glyphs to name but a few. It has only two floating marks; all other symbols are normal spacing characters. The alphabetic order of the consonants is similar to the typical Indic order.

Tai Mau is written from left to right (with spaces between words? syllables?). Each syllable begins with a consonant (or glottal stop?) followed by a vowel, any final stop follows the vowel, and finally comes a tone mark. Tone marks are spacing characters; the first tone is indicated by absence of any other tone mark. There are no special symbols for final consonants: consonants are known to be final stops by virtue of their position within a syllable after a vowel, since all vowels are explicitly marked. (is that strictly true?). As in the Indic systems, the consonants also contain an inherent conceptual vowel. This inherent vowel in Tai Mau represents both the vowel "a" and a glottal stop. To write the vowel "a" without glottal stop, a special symbol (like a lowercase 'b') is used.

Foreign sounds are expressed principally through use of a non-spacing dot. This dot may be written either on the upper right shoulder of a vowel, or below the vowel, to shorten its value. Placing the dot over the *tone* symbol indicates a rising tone; and placing it below the tone symbol indicates a falling tone. Voiced consonants are written by applying the dot under a consonant (e.g., to turn 'k' into 'g'). More than one final stop may be written by putting a dot above the 2nd (and nth) final consonants of a syllable.

Issues: Several issues are framed as questions in the paragraphs above. The script seems, from the available sources, to be deceptively simple. It is not known at all how widely this system is currently used, but it is assuredly in modern use.

Punctuation and word spacing and so forth are currently unknown.

There are some diphthongs that are written with combinations of primitive vowel signs followed by ''sha1'', and some diphthongs written with combinations of primitive vowel signs followed by *what appears to be* the consonant WA. The diphthong listed as ''ai bar'' in the names list is written with a unique symbol that looks like the vowel sign AA, but has the hook to the right; it is not clear whether this is an error in the source or not.

There is no ''tone mark 1'' in the chart or names list since the unmarked state is what we shall call tone 1.

Some Sources

Young, Linda Wai Ling. *Shan Chrestomathy.*

Rev 92/11/25

Tai Mau, draft names list, 92/10/21

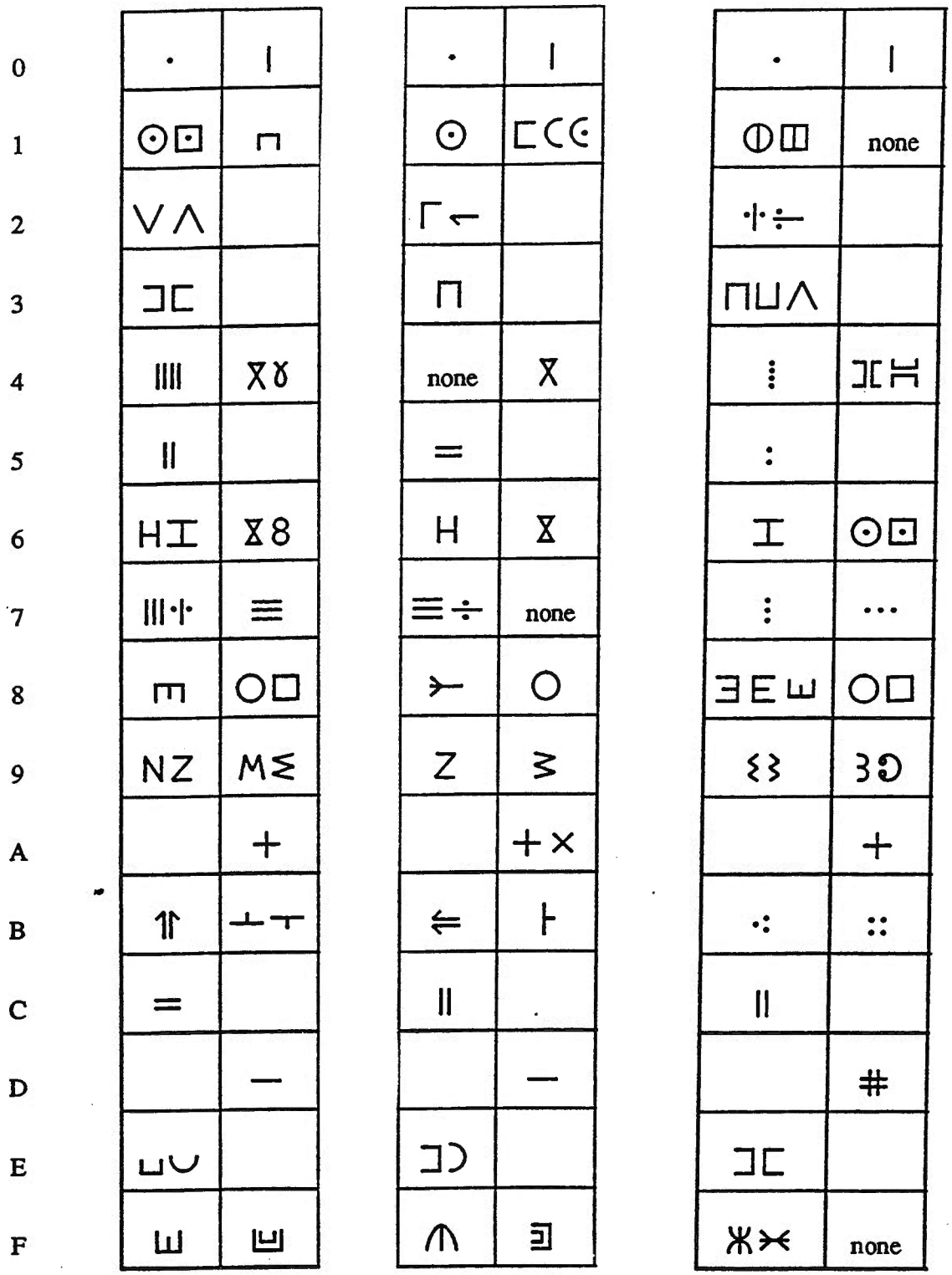
00 TAI MAU LETTER KA
01 TAI MAU LETTER KHA
02 TAI MAU LETTER NGA
03 TAI MAU LETTER TSA
04 TAI MAU LETTER SA
05 TAI MAU LETTER NYA
06 TAI MAU LETTER TA
07 TAI MAU LETTER THA
08 TAI MAU LETTER NA
09 TAI MAU LETTER PA
0A TAI MAU LETTER PHA
0B TAI MAU LETTER FA
0C TAI MAU LETTER MA
0D TAI MAU LETTER YA
0E TAI MAU LETTER RA
0F TAI MAU LETTER LA

10 TAI MAU LETTER WA
11 TAI MAU LETTER HA
12 TAI MAU LETTER AH
13 TAI MAU LETTER SHA1
14 TAI MAU LETTER SHAA
15 TAI MAU LETTER SHA2
16 TAI MAU TONE MARK 2
17 TAI MAU TONE MARK 3
18 TAI MAU TONE MARK 4
19 TAI MAU TONE MARK 5
1A TAI MAU TONE MARK 6
1B TAI MAU VOWEL SIGN A
1C TAI MAU VOWEL SIGN AA
1D TAI MAU VOWEL SIGN I
1E TAI MAU VOWEL SIGN E
1F TAI MAU VOWEL SIGN EE

20 TAI MAU VOWEL SIGN U
21 TAI MAU VOWEL SIGN O
22 TAI MAU VOWEL SIGN OH
23 TAI MAU VOWEL SIGN I BAR
24 TAI MAU VOWEL SIGN SCHWA
25 TAI MAU VOWEL SIGN AI BAR
26 TAI MAU FALLING TONE OR VOICE MARK
27 TAI MAU RISING TONE OR SHORT VOWEL

NUMIDIAN / TIFINAGH

92/10/29
(92/2/2)



0	·	ǀ
1	⊙⊡	⊓
2	⋁⋀	
3	⊐⊏	
4	ⅡⅡ	ⵅⵖ
5	‖	
6	H⌶	ⵅ8
7	Ⅲ⁘	≡
8	⊓	○□
9	NZ	M≤
A		+
B	⇑	⊥⊤
C	=	
D		—
E	⊔∪	
F	⊔	⊔

NUMIDIAN

0	·	ǀ
1	⊙	⊏⊂⊂
2	⌜←	
3	⊓	
4	none	ⵅ
5	=	
6	H	ⵅ
7	≡÷	none
8	⊁	○
9	Z	≤
A		+×
B	⇐	⊦
C	‖	
D		—
E	⊐⊃	
F	⋀	⊐

TIFINAGH

0	·	ǀ
1	ⵔⵕ	none
2	⁜÷	
3	⊓⊔⋀	
4	⁝	⊐⊏
5	:	
6	⌶	⊙⊡
7	⁝	⋯
8	ⵎⵉⵡ	○□
9	ⵙⵣ	ⵣⵒ
A		+
B	⁖	⁘
C	‖	
D		#
E	⊐⊏	
F	ⵟⵠ	none

Tifinagh, Numidian

Tifinagh is a living script used among the Berber people of the Sahara. It seems to be a direct descendant of the ancient Numidian script, with which it shares many of its letter forms. (Numidian is also called *Libyan* by Diringer who notes that it is contemporaneous with the Roman period.) Unfortunately, not much more is known about it at this time. It was apparently influenced by Punic.

Numidian was normally written from bottom to top, in columns from left to right. In some bilingual Numidian and Punic inscriptions, the Numidian parts were written from right to left horizontally in the Punic manner.

Modern Tifinagh is apparently written horizontally, from right to left with lines running from top to bottom. There are some ligatures used in writing Tifinagh. It is not known whether they are obligatory or not in Tifinagh rendering.

Neither Tifinagh nor Numidian uses any diacritical marks or other non-spacing characters. Some of the glyphs in both Numidian and Tifinagh change form depending on whether they are being written horizontally or vertically.

Issues: The script called *Tamachek* may be the same thing as Tifinagh. The names list is purely for identification and must be revised when information becomes available.

It is not at all clear whether Tifinagh should be encoded separately from Numidian or whether they should be encoded as a single composite script. Some of the graphic elements used for one phonetic value in Tifinagh were used for a completely different phonetic value in Numidian. Fairly solid information on Tifinagh, including ligatures and the alphabet, is currently available, as is information on Numidian. Since they have very high overlap in terms of signs, it seems reasonable to encode them either in parallel or as a single script, depending primarily upon graphic form for the choice of the character complement. Not enough information is available about the history of either to make this proposal very complete.

The accompanying charts were prepared from draft charts supplied by Lloyd Anderson. They are laid out to match each other phonetically, and are both parallel to the Unicode Hebrew block. They are here supplied together for

information and comparison. The left hand group is Numidian, with glyphs for *vertical* writing. The middle group is Numidian, with glyphs for *horizontal* writing. The right hand group is modern Tifinagh.

Some Sources

Friedrich, Johannes. *Extinct Languages.*
Diringer, David. *Writing.*

Rev 92/10/23

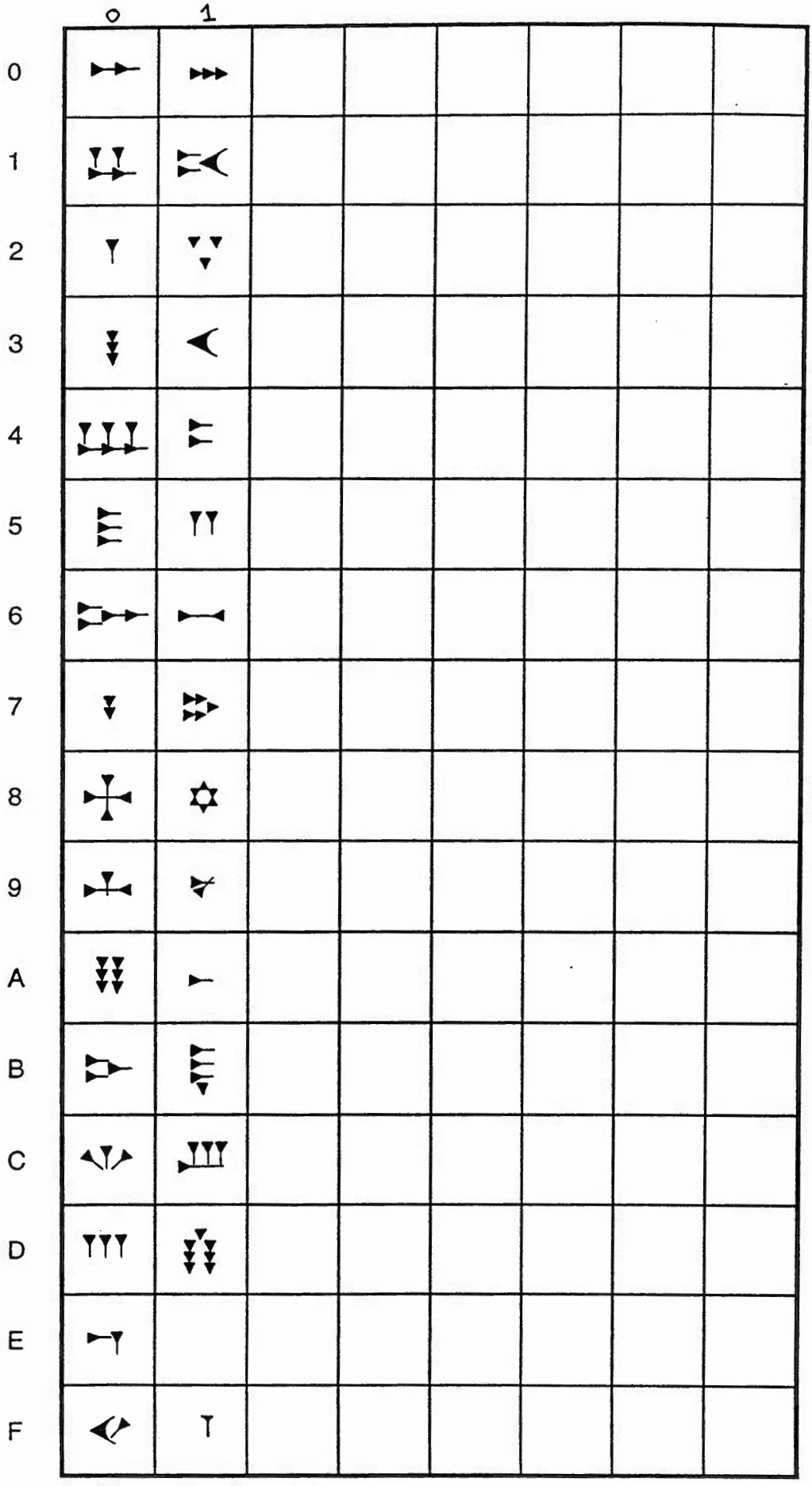
Numidian Names draft, 92/10/23 (parallel to Hebrew)

00 NUMIDIAN LETTER ALPHA
01 NUMIDIAN LETTER B
02 NUMIDIAN LETTER G HACEK
03 NUMIDIAN LETTER D
04 NUMIDIAN LETTER H
05 NUMIDIAN LETTER U UNDERBAR
06 NUMIDIAN LETTER Z HACEK
07 NUMIDIAN LETTER G OVERDOT
08 NUMIDIAN LETTER T UNDERDOT
09 NUMIDIAN LETTER I UNDERBAR
0A
0B NUMIDIAN LETTER K
0C NUMIDIAN LETTER L
0D
0E NUMIDIAN LETTER M
0F NUMIDIAN LETTER Z OVERBAR

10 NUMIDIAN LETTER N
11 NUMIDIAN LETTER S TWO
12
13
14 NUMIDIAN LETTER P (F)
15
16 NUMIDIAN LETTER S
17 NUMIDIAN LETTER Q
18 NUMIDIAN LETTER R
19 NUMIDIAN LETTER S HACEK
1A NUMIDIAN LETTER T
1B NUMIDIAN LETTER H UNDERBAR
1C
1D NUMIDIAN LETTER Z
1E
1F NUMIDIAN LETTER T TWO

Ugaritic

draft
DEC. 10, 1991



Ugaritic Cuneiform

The city state of Ugarit was an important seaport on the Phoenician coast (directly east of Cyprus, north of the modern town of Minet el-Beida) from about 1400 BC until it was completely destroyed in the 12th century BC. The site of Ugarit, now called Ras esh-Shamra, was apparently continuously occupied from Neolithic times (ca. 5000 BC). It was first uncovered by a local inhabitant while ploughing a field in 1928, and subsequently excavated by Claude Schaeffer and Georges Chenet beginning in 1929, in which year the first of many tablets written in the Ugaritic script were discovered. They later proved to contain extensive portions of an important Canaanite mythological and religious literature that had long been sought and which revolutionized Biblical studies. The script was first deciphered in a remarkably short time jointly by Hans Bauer, Édouard Dhorme, and Charles Virolleaud.

The Ugaritic language is Semitic, variously regarded by scholars as being a distinct language related to Akkadian and Canaanite, or a Canaanite dialect. Ugaritic is generally written from left to right horizontally, sometimes with a vertical stroke between words. In the city of Ugarit, this script was also used to write the Hurrian language.

Glyphs for *T-Underbar*, *G-Acute*, and *D-Underbar* differ somewhat between modern reference sources (as do some transliterations). T-Underbar is most often displayed with a glyph that looks like an occurrence of *Glottal Stop* overlaid with *G*. The Unicode block for Ugaritic is in the order that was apparently standard; it coincides for the mostpart with Phoenician and Early Hebrew order.

Ugaritic cuneiform is thought to be complete in this encoding; it is a syllabic script and should not be confused with the ideographic cuneiform scripts of Akkadian and Sumerian derivation. There may be relatives of the Ugaritic script used for other Canaanite languages at about the same time.

Issues: Because the Ugaritic language was Semitic, and therefore the script contains syllables which somewhat echo the Semitic alphabets, it has been suggested that scholars could benefit were it to be encoded in phonetic parallel to the Hebrew script.

Some Sources

Cleator, P. E. *Lost Languages.*
Coulmas, Florian. *Writing Systems of the World.*
Friedrich, Johannes. *Extinct Languages.*
Gordon, Cyrus H. *Forgotten Scripts.*

Rev 92/10/20

Ugaritic Names List, draft 92/10/29

00 UGARITIC LETTER A
01 UGARITIC LETTER B
02 UGARITIC LETTER G
03 UGARITIC LETTER H UNDERBAR
04 UGARITIC LETTER D
05 UGARITIC LETTER H
06 UGARITIC LETTER W
07 UGARITIC LETTER Z
08 UGARITIC LETTER H UNDERDOT
09 UGARITIC LETTER T UNDERDOT
0A UGARITIC LETTER Y
0B UGARITIC LETTER K
0C UGARITIC LETTER S BREVE
0D UGARITIC LETTER L
0E UGARITIC LETTER M
0F UGARITIC LETTER D UNDERBAR

10 UGARITIC LETTER N
11 UGARITIC LETTER T UNDERBAR UNDERDOT
12 UGARITIC LETTER S
13 UGARITIC LETTER GLOTTAL STOP (ain)
14 UGARITIC LETTER P
15 UGARITIC LETTER S UNDERDOT
16 UGARITIC LETTER Q
17 UGARITIC LETTER R
18 UGARITIC LETTER T UNDERBAR
19 UGARITIC LETTER G ACUTE
1A UGARITIC LETTER T
1B UGARITIC LETTER I
1C UGARITIC LETTER U
1D UGARITIC LETTER S GRAVE
1E
1F UGARITIC WORD DIVIDER

Other Scripts (Without Specific Proposals)

There are, of course, a number of other scripts for which proposals have not been made. Some of these will be described in this section. Further information about these scripts is welcome. Scholars interested in pursuing the encoding of any of these may contact the Unicode offices. In the following thumbnail sketches, when it is written that a particular item "is not known," this usually means that the relevant information has not yet been found by members of the Unicode Consortium working on these issues, rather than that the information is really not known.

Brahmi and Other Scripts of India

The Brahmi script is the progenitor of all or most of the scripts of India, as well as most scripts of Southeast Asia. Brahmi is also known as *Asoka*, the script in which the famous Asokan edicts were incised in the second century BC. (Asoka was an emperor of the Mauryan dynasty of what is now Orissa State, India.) Brahmi is historically important, but not enough information is currently available to make a concrete proposal beyond a mere list of the basic alphabet (e.g., for which see Diringer's *Writing*). Unlike most of its modern descendants, Brahmi vowel signs are written in an attached form, and the script thus requires a large number of glyphs for rendering.

The so-called **Box-Headed Script** was used in India during the 6th century AD. It appears in many stone inscriptions around Hyderabad in central India. Several other old Indian scripts are known to exist (**Modi, Kaithi, Satavahana, Chola, Kharoshthi, Lahnda**) but not enough information is currently available about them to evaluate their content and historical importance. They may eventually be encoded.

'Phags-pa

The 'Phags-pa script an extinct fore-runner of the Tibetan script, is traditionally held to have been invented in about 1269 by *Bla-ma 'Phags-pa*. It was used in Mongolia throughout the Yüan dynasty and (reportedly) was the official script of the Mongolian empire under Kublai Khan. 'Phags-pa can be viewed as mostly parallel to the modern Tibetan script, but it was written vertically and contained several letters not found in Tibetan.

Ancient Egyptian (Hieroglyphic)

The Egyptian hieroglyphic script is well-known and historically important; it is also well-studied by scholars and frequently requested for addition to Unicode. The major problem to solve is determining the extent to which variant forms should be unified into a single codepoint, relying on richer text handling mechanisms for rendering and glyphic choice. The Gardner set of glyphs contains some 750 entities from a late hieroglyphic period. French scholars have compiled some 9000 entities spanning from the earliest to latest inscriptions; of these 9000, one preliminary estimate suggests that only about 2000 should really be distinct characters, the other 7000 are variant forms. A clear model needs to be developed that can give a coherent picture of the historical periods involved, and how various periods can be reflected in the final rendering and processing models. So far, no work has been done in this area. This problem is of similar magnitude to the "Han unification" problem.

Akkadian / Babylonian / Sumerian

The Egyptian hieroglyphic problem is probably closely matched by the problems involved in the **Akkadian**, **Sumerian**, and **Babylonian** cuneiform systems. One existing Akkadian font lists over 700 signs. The *Manuel d'Epigraphie Akkadienne* has not been available for preliminary consultation (it has been purchased, but we have yet to receive it as of this writing). Akkadian was a *lingua franca* over much of the ancient Middle East for well over a thousand years, and its historical importance is uncontested, but again, there is an historical problem of considerable magnitude to be solved before encoding it.

Hittite Hieroglyphics

The Hittite language written with a unique hieroglyphic system is the oldest recorded Indo-European language. The Hittite hieroglyphics came to light gradually during the latter half of the 19th century. There are some 110 signs or so. Many of these are listed in various readily-available sources, but we have not yet found source materials showing *all* of the known signs or expounding more than cursorily upon the hieroglyphic system. Hittite was also written at one time in a later form of Akkadian cuneiform; it is not known to what extent the glyphs used for cuneiform Hittite overlap exactly with particular Akkadian glyphs.

Kawi / Javanese / Balinese

It is not clear at this time whether Kawi, Javanese, and Balinese scripts are distinct enough entities to require separate encoding, or whether a single encoding with three different font presentations will suffice. The Javanese script is known to enjoy some sporadic use, and some information on it (the shapes and phonetic values of its basic letters, from Faulmann and other sources) is readily available. Kawi is basically an extinct language, but it is known to still enjoy some use at least in traditional Balinese theatre (see, e.g., McPhee, *Music in Bali* – where the Kawi language is mentioned repeatedly as the language of vocal recitation for much theatre music). It is unknown to what extent either the Kawi or Balinese scripts are in use, however.

Ahom / Khamti

Ahom is a recently extinct Shan language. The Ahom and Khamti scripts appear in the *Linguistic Survey of India* (see below), where there is enough information to quickly generate an exploratory proposal. Hearsay suggests, however, that a new book on the Ahom language (and script?) is forthcoming; this could be expected to contain much better information. It is unknown how much current scholarly interest there is in encoding of the Ahom and Khamti scripts.

Pyu / Tircul

The Pyu script is another descendant of Brahmi that was used in Burma sometime between about 800 and 1000 AD. It is described somewhat in Luce (see below), where there is a large chart that gives a good idea of the letter shapes and the repertoire, but is too scanty for even an exploratory proposal.

Yi (Lolo)

The Yi or Lolo script is known to be in use among the Yi people of Yunnan Province in China. The modern Yi script is a syllabary containing hundreds of symbols. Each symbol seems to encode a syllable and one of three tones. A table of this script is available. The system seems to be a revision of an older syllabic/ideographic system about which little information is available. Some further other information is contained in Vial (see below).

Moso (a.k.a. Naxi, Nahsi, Nakhi)

The Moso or Naxi script is used among the Moso people of China. It is apparently an ideographic script (with many beautiful and detailed glyphs), and may still be in use as of this writing. It was apparently in use as late as 1981. Bacot shows a large number of ideographs, with brief synopses of meaning. This information is adequate to get an idea of the number of symbols and their type, but more information is needed to generate an exploratory proposal. One volume in Chinese (1981) is available and lists some 1340 graphic units, though this number must be augmented because several dissimilar graphic elements are often recorded and defined under one numbered entry.

Siddham

The Siddham script is closely related to Devanagari. It is still widely used as an art form (calligraphy) in connection with Buddhism in Japan and the Far East. Excellent sources, such as Stevens (see bibliography) are available, and a proposal could be quickly generated.

Linear A and Others

Several other scripts are known from the Middle East. Among these are **Linear A** and the **Cypriot Syllabary** (or Cypro-Minoan). They are both related to Linear B but the extent of the connection is not clear enough to decide whether they could or should be encoded in parallel to Linear B or unified with Linear B, or encoded separately. Not much information is available on the so-called "pseudo-hieroglyphic" script of **Byblos**.

Some Sources

Luce, G. H. *Phases of Pre-Pagán Burma.*
Vial, Paul. *Les Lolos.*
Bacot, J. *Les Mo-so.*
Grierson, G. A. *Linguistic Survey of India.*
Gordon, Cyrus H. *Forgotten Scripts.*
Stevens, John. *Sacred Calligraphy of the East.*

Bibliography

Alexander, J. T. *A Dictionary of the Cherokee Indian Language.* Published by the author, 1971.

Antonsen, Elmer H. *The Runes: The Earliest Germanic Writing System*, in The Origins of Writing. Wayne M. Senner, ed. Univ. of Nebraska Press, Lincoln, 1989.

Bacot, J. *Les Mo-so; Ethnographie de Mo-so, leurs religions, leur langue et leur écriture.* E. J. Brill, Leide, 1913.

Bonfante, Larissa. *Etruscan.* University of California Press / British Museum, Berkeley, 1990. *Reading the Past* Series.

Budge, E. A. Wallis. *The Rosetta Stone.* Dover. New York, 1989. ISBN 0-486-26163-8 [First published 1929].

Campbell, A. *Note on the Limboo Alphabet of the Sikkim Himalaya* in Journal of the Asiatic Society of Bengal, Vol 24, 1855.

Chadwick, John. *Linear B and Related Scripts.* University of California Press / British Museum, Berkeley, 1987. *Reading the Past* Series.

Chemsong, Iman Singh. *The Kirat Grammar (Limbu).* PL3801.L91C5 Information incomplete.

Cleator, P. E. *Lost Languages.* The John Day Co. New York, 1961. LC 61-8278.

Cook, B. F. *Greek Inscriptions.* University of California Press / British Museum, Berkeley, 1987. *Reading the Past* Series.

Coulmas, Florian. *Writing Systems of the World.* Basil Blackwell, Oxford, 1989.

Cross, Frank Moore. *The Invention and Development of the Alphabet*, in The Origins of Writing. Wayne M. Senner, ed. Univ. of Nebraska Press, Lincoln, 1989.

Davies, W. V. *Egyptian Hieroglyphs.* University of California Press / British Museum, Berkeley, 1990. *Reading the Past* Series.

Davis, Richard. *A Northern Thai Reader.* The Siam Society, Bangkok, 1970.

Diringer, David. *Writing.* Frederick A. Praeger Publisher, New York, 1962.

Diringer, David. *The Story of the Aleph Beth.* Thomas Yoseloff, New York, 1960.

Encyclopaedia Brittanica, 15th edition (1981), Articles: *Anatolian languages, Ancient epigraphic remains, Alphabets, Etruscan language Luwian, Lycian alphabet, Lycian language, Lydian language.*

Faulmann, Carl. *Schriftzeichen und Alphabete aller Zeiten und Völker.* Augustus Verlag, Augsburg, 1990. Reprint of 1880 edition.

Fossey, Charles. Notices sur les caract©ères ©étrangers, anciens Impr. nationale de France, Paris, 1948.

Francisco, Juan R. *Philippine Palaeography.* Philippine Journal of Linguistics, Special Monograph Issue Number 3. Linguistic Society of the Philippines, Quezon City, 1973.

Friedrich, Johannes. *Extinct Languages.* Philosophical Library, New York, 1957. (Translation of *Entzifferung Verschollener Schriften und Sprachen.*)

Gardiner, A. H. *Egyptian Grammar.* London, 1957. [Reprinted by Dover?]

Gelb, I. J. *Hittite Hieroglyphics, I, II, III.* Chicago, 1931, 1935. [Not found for consultation.]

Gordon, Cyrus H. *Forgotten Scripts.* Basic Books, New York, 1968.

Gordon, Cyrus H. *Ugaritic Literature.* Ventnor Publishers, Ventnor NJ, 1949. [Not found for consultation. Cited as source of Cleator's Ugaritic table.]

Gorer, Geoffrey. *Himalayan Village, an account of The Lepchas of Sikkim.* Second Edition. Basic Books, New York, 1967. (First pub. London, 1938).

Graves, Robert. *The White Goddess: a historical grammar of poetic myth.* Noonday Press, New York, 1948 (1990 reprint).

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

Haarh, Erik. *The Lepcha Script*, in Acta Orientalia 24, 1959, pp 107-122.

Haug, Martin & Destur Hoshangji Jamaspji Asa. *An Old Pahlavi-Pazand Glossary.* Biblio Verlag, Osnabrück, 1973. (Reprint of 1870 edition.)

Holmes, Ruth Bradley & Betty Sharp Smith. *Beginning Cherokee.* University of Oklahoma Press, Norman. [Publication date unknown.]

Healey, John F. *The Early Alphabet.* University of California Press / British Museum, Berkeley, 1990. *Reading the Past* Series.

Haugen, Einar. *History of the Scandinavian Language(s).* Information incomplete.

Jackson, A. V. Williams. *An Avesta Grammar in Comparison with Sanskrit.* Part 1, Phonology, Inflection, Word-Formation. AMS Press, 1975. (Reprint of the 1892 edition of W. Kohlhammer, Stuttgart.)

Kilpatrick, Jack Frederick & Anna Gritts Kilpatrick, eds. *New Echota Letters: Contributions of Samual A. Worcester to the Cherokee Phoenix.* Southern Methodist Univ. Press, Dallas, n.d. (Reprint of an article by S. A. Worcester which appeared in the Cherokee Phoenix, Feb. 21, 1828).

Kirat Primary Book. 1970. Information incomplete.

Lehmann, Ruth P. M. *Ogham: Ancient Script of the Celts*, in The Origins of Writing. Wayne M. Senner, ed. Univ. of Nebraska Press, Lincoln, 1989.

Library of Congress. *Cataloging Service Bulletin*, No. 19 / Winter 1982.

Limbu Reader VI. LC 82-90304. Information incomplete.

Luce, G. H. *Phases of Pre-Pagán Burma; Languages and History.* Oxford University Press, Oxford, 1985. [Pyu, Tircul]

MacKenzie, D. N. *A Concise Pahlavi Dictionary.* Oxford University Press, London, 1971.

Mainwaring, G. B. *A Grammar of the Róng (Lepcha) Language.* Printed by C. B. Lewis, Baptist Mission Press, Calcutta, 1876. (Recently reprinted by Ratna Pustak Bandhar, Kathmandu.)

Mainwaring, G. B. *Dictionary of the Lepcha Language.* Revised and completed by A. Grünwedel, Berlin, 1898.

McPhee, Colin *Music in Bali*, Yale Univ Press, New Haven, 1966

Nakanishi, Akira. *Writing Systems of the World.* Tuttle. Rutland, VT, 1980. (Translation of *Sekai no moji*, Shokado, Kyoto, 1975.) ISBN 0-8048-1293-4. LC 79-64826.

Nakano, Miyoko. *A Phonological Study in the 'Phags-pa Script and the Meng-ku Tzu-yün.* Faculty of Asian Studies in association with Australian National University Press, Canberra, 1971.

Norman, James. *Ancestral Voices; decoding ancient languages.* Four Winds Press, New York, 1975.

Nyberg, Henrik Samuel. *A Manual of Pahlavi.* Otto Harrassowitz, Wiesbaden, 1964. Second edition of *Hilfsbuch des Pehlevi.*

Page, R. I. *Runes* (University of California Press / British Museum, Berkeley, 1990). *Reading the Past* Series.

Pontalis, Pierre Lefevre. *L'invasion Thaïe en Indo-Chine*, in T'oung pao Archives, Vol VIII. E. J. Brill, Leide, 1897. Kraus Reprint, Nendeln, Liechtenstein, 1975.

Sampson, Geoffrey. *Writing Systems; a linguistic introduction.* Stanford University Press, Stanford, CA, 1985.

Senner, Wayne M., ed. *The Origins of Writing*. University of Nebraska Press, Lincoln, 1989. [Several articles also cited.]

Sirk, Ü. *The Buginese Language.* Nauka Publishing House, Central Department of Oriental Literature, Moscow, 1983. *Languages of Asia and Africa* series.

Sloat, Clarence & Sharon Henderson Taylor & James E. Hoard *Introduction to Phonology.* Prentice Hall, Englewood Cliffs, 1978. [Cherokee table.]

Stevens, John. *Sacred Calligraphy of the East.* Shambala. Boston, 1988. [Source for *Siddham* script.]

Subba, B. B. *Limbu Nepali English Dictionary.* Gangtok, Sikkim, 1979. PL3801.L54S9 1979.

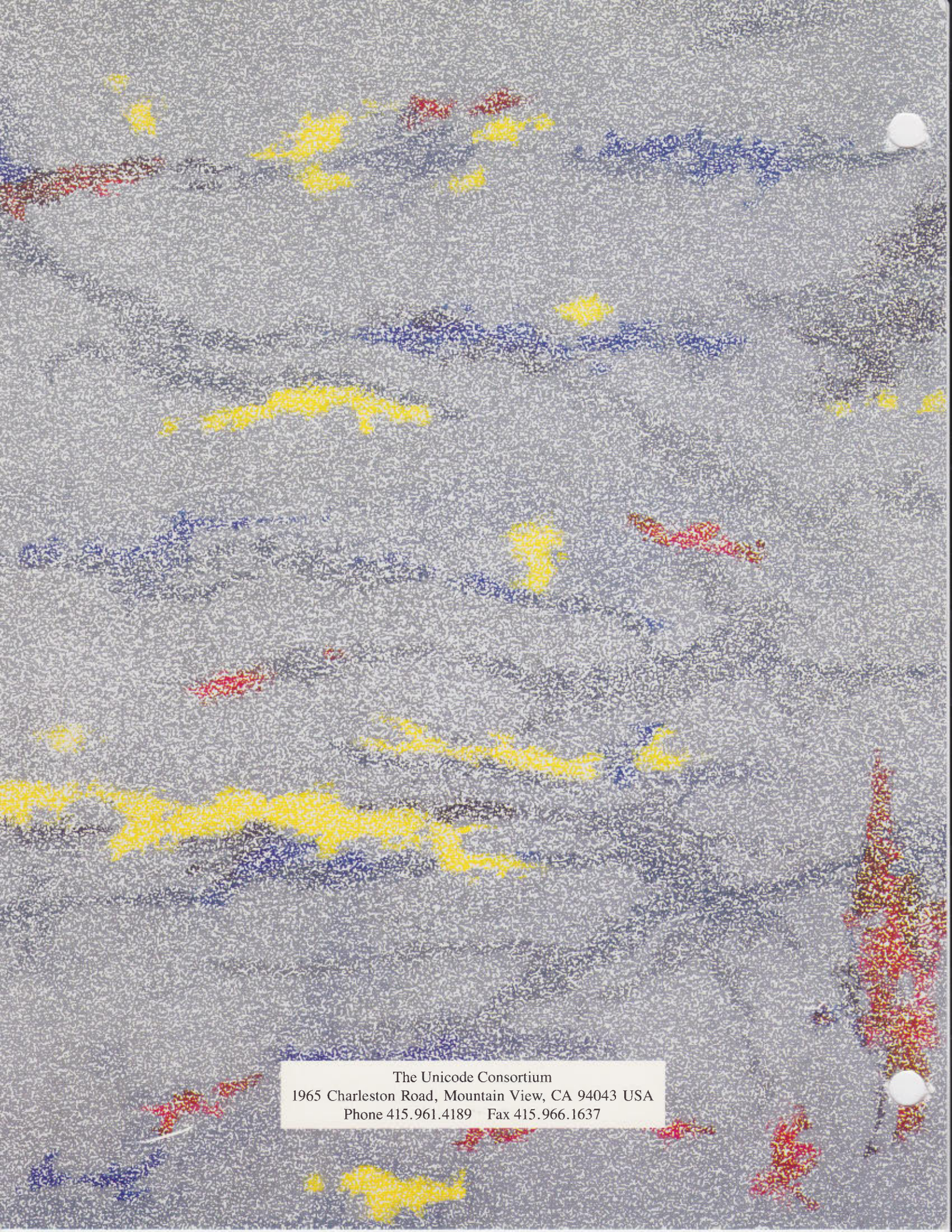
van der Tuuk, H. N. *A Grammar of Toba Batak.* Martinus Nijhoff, The Hague, 1971. (Translation of 1864 work.)

Vial, Paul. *Les Lolos; histoire, religion, moeurs, langue.* Chang-Hai, Imprimerie de la Mission Catholique, 1898.

Walker, C. B. F. *Cuneiform.* University of California Press / British Museum, Berkeley, 1987. *Reading the Past* Series.

Xerox Character Code Standard. Xerox System Integration Standard XNSS 059003, June 1990, Version 2.0.

Young, Linda Wai Ling. *Shan Chrestomathy; an introduction to Tai Mau Language and Literature.* Monograph series no. 28. Center for South and Southeast Asia Studies, University of California, Berkeley, 1985.



The Unicode Consortium
1965 Charleston Road, Mountain View, CA 94043 USA
Phone 415.961.4189 Fax 415.966.1637