JTC1/SC2/WG2 N3645R
L2/09-187R
2009-07-28

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

Doc Type: Working Group Document
Title: Proposal for encoding the Chakma script in the UCS
Source: UC Berkeley Script Encoding Initiative (Universal Scripts Project)
Authors: Michael Everson and Martin Hosken
Status: Liaison Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2009-07-28
Replaces: N3428

1. Introduction. The Chakma people were in origin Tibeto-Burman, related to the Burmese. The language which they now speak is Indo-European, part of the Southeastern Bengali branch of Eastern Indo-Aryan. Its better-known closest relatives are Bengali, Assamese, Chittagonian, Bishnupriya, and Sylheti. It is spoken by 312,000 people in southeast Bangladesh near Chittagong City, and another 176,000 in India in Mizoram, Assam, Tripura, and Arunachal Pradesh. Literacy in Chakma script is low. The script itself is also called 𑄃𑄏 𑄛𑄒𑄴 *Ajhā pāṭh*, sometimes romanized *Ojhopath*.

There is a certain amount of glyph variation between the script as used in India and Bangladesh. Some fonts are rounder, similar to the style used in Myanmar; compare a similar variation in the Tai Tham script as used in Khün Tai. The glyphs used in this proposal are based on the Chadigang font, with some alterations toward more "generic" shapes for some characters.

The Chakma script is currently being adapted for use in Tanchangya, a language which is closely related to Chakma. An effort to develop the orthography is currently underway, and it appears that there may be additional letters, vowel signs, and tone marks added to cover this script. These extensions are a subject for future standardization, as the orthography for Tanchangya is still under development and testing.

2. Structure. Chakma is of the Brahmic type: the consonant letters contain an inherent vowel. Consonant clusters are written with conjunct characters, and a visible vowel killer shows the deletion of the inherent vowel when there is no conjunct.

3. Independent vowels. Four independent vowels exist: 𑄃 *a*, 𑄄 *i*, 𑄅 *u*, and 𑄆 *e*. Other vowels in initial position are formed by adding the vowel sign to 𑄃 *a*, as in 𑄃𑄩 *ī*, 𑄃𑄫 *ū*, 𑄃𑄭 *ai*, 𑄃𑄰 *oi*. Some modern writers are generalizing this spelling in 𑄃𑄨 *i*, 𑄃𑄪 *u*, and 𑄃𑄬 *e*.

4. Dependent vowels. Independent vowel signs have been encoded according to their phonetic value, as in Balinese or Telugu. However, because some of the independent vowels appear to be made out of smaller units, decomposition of the vowels has been taken into consideration. In order to avoid any ambiguous encoding, canonical equivalences have been given in order to resolve the issue of multiple representations. Two glyph fragments (CHAKMA O MARK and CHAKMA AU MARK) have been encoded in order to account for this, and to allow for any users who desire to create texts using the lower glyph fragments on their own. The Chakma vowel signs are given with the letter 𑄇 *ka* below.

𑄇 kā = 𑄇 kā

𑄈 ka = 𑄇 kā + ◌𑄴 -a (11127)

𑄉 ki = 𑄇 kā + ◌𑄰 i (11128)

𑄊 kī = 𑄇 kā + ◌𑄱 -ī (11129)

𑄋 ku = 𑄇 kā + ◌𑄲 -u (1112A)

𑄌 kū = 𑄇 kā + ◌𑄳 -ū (1112B)

𑄍𑄇 ke = 𑄇 kā + 𑄍◌ -e (1112C)

𑄎 kāi = 𑄇 kā + ◌𑄵 -āi (1112D)

𑄏 ko = 𑄇 kā + ◌𑄴𑄧 -o (1112E – *or* – 11131 + 11127)

𑄐 kau = 𑄇 kā + ◌𑄴𑄱 -au (1112F – *or* – 11132 + 11127)

𑄑 koi = 𑄇 kā + ◌𑄶 -oi (11130)

𑄀 kaṁ = 𑄇 kā + ◌𑄀 -ṁ (11100)

𑄁 kaṃ = 𑄇 kā + ◌𑄁 -ṃ (11101)

𑄂 kaḥ = 𑄇 kā + ◌𑄂 -ḥ (11102)

𑄃 k = 𑄇 kā + ◌𑄃 MAAYYAA

One of the interesting features of Chakma writing is that CANDRABINDU (*cānaphupudā*) can be used together with ANUSVARA (*ekaphudā*) and VISARGA (*dviphudā*):

𑄅𑄀 aḥṁ = 𑄅 ā + ◌𑄂 ḥ + ◌𑄀 ṁ

𑄅𑄀 aṃṁ = 𑄅 ā + ◌𑄁 ṃ + ◌𑄀 ṁ

𑄆𑄀 uṃṁ = 𑄆 u + ◌𑄁 ṃ + ◌𑄀 ṁ

𑄌𑄀 muṁ = 𑄉 mā + ◌𑄲 u + ◌𑄀 ṁ

Glyph rendering of CANDRABINDU with other vowel signs (like ◌𑄴 *a* or ◌𑄰 *i*) is as yet unattested; nevertheless the encoding is clear.

5. Consonants with killed vowels and conjunct consonants. Like other Brahmic scripts, Chakma makes use of the MAAYYAA (*killer*) to invoke conjoined consonants. In the past, practice was much more common than it is today. Like the Myanmar script, Chakma is encoded with two vowel-killing characters in order to conform to modern user expectations. As shown above, most letters have their vowels killed with the use of the explicit MAAYYAA character:

𑄃 k = 𑄇 kā + ◌𑄃 MAAYYAA

In 2001 an orthographic reform was recommended in the book *Cāṅmā pattham pāt* which would limit the standard repertoire of conjuncts to those composed with the five letters 𑄝 *yā*, 𑄞 *rā*, 𑄟 *lā*, 𑄠 *wā*, and 𑄡 *nā*. The four here are the most widely-accepted repertoire of conjuncts.

ya: X + 𑄳 VIRAMA + 𑄪 *yā*:

ra: X + 𑄳 VIRAMA + 𑄨 *rā*:

la: X + 𑄳 VIRAMA + 𑄩 *lā*:

wa: X + 𑄳 VIRAMA + 𑄫 *wā*:

No separate conjunct forms of subjoined full-form *-yā* or *-rā* appear to exist. The fifth of these conjuncts, the *-na* conjunct, is exemplary of the orthographic shift which has taken place in Chakma.

na: X + 𑄳 VIRAMA + 𑄝 *nā*:

While some writers would indeed write *kakna* (in ligating style) as 𑄌𑄝 or (in subjoining style) as 𑄌𑄝𑄳, most now would probably expect it to be written as 𑄌𑄳𑄝. The ligating style of glyphs is now considered old-fashioned. Thus, taking the letter 𑄪 *mā* as the second element, while the glyph shapes 𑄌𑄪 *kmā*, 𑄓𑄪 *tmā*, 𑄝𑄪 *nmā*, 𑄚𑄪 *bbā*, 𑄟𑄪 *mmā*, 𑄢𑄪 *llā*, 𑄞𑄪 *smā*, and 𑄠𑄪 *hmā* are attested, most users now prefer the glyph shapes 𑄌𑄴 *kmā*, 𑄓𑄴 *tmā*, 𑄝𑄴 *nmā*, 𑄚𑄴 *bbā*, 𑄟𑄴 *mmā*, 𑄢𑄴 *llā*, 𑄞𑄴 *smā*, and 𑄠𑄴 *hmā*. Again, this distinction is stylistic and not orthographic.

As with Myanmar and Meetei Mayek, encoding a visible killer for modern users alongside an explicit conjoin-former permits the user to make specific choices about spelling more easily. Both the Myanmar encoding model and the Devanagari encoding model have been explained to the user community and feedback is that the Myanmar model fits the script better. (This is little surprise considering the close relationship between the Myanmar and Chakma scripts.)

In principle, nothing prevents the visible killer from appearing together with a stack. Thus while 𑄌 *kā* + 𑄳 VIRAMA + 𑄞 *sā* gives 𑄌𑄳 *ksā*, the sequence 𑄌 *kā* + 𑄳 VIRAMA + 𑄞 *sā* + 𑄧 MAAYYAA could give 𑄌𑄳 *ks*. Neither the string 𑄳 VIRAMA + 𑄧 MAAYYAA nor the string 𑄧 MAAYYAA + 𑄳 VIRAMA following a consonant would be meaningful, however, as both kill the inherent vowel, and “double-killing” a vowel makes no sense; either sequence should generate an error presentation.

The 2004 book *Phadagaṅ* shows examples of the five conjuncts above together alongside conjuncts formed with 𑄚 *bā*, 𑄪 *mā*, and 𑄠 *hā*. These are all formed by simple subjoining.

ba: X + ◌್ VIRAMA + ಬ bā:

ಕ್ಬ ಖ್ಬ ಗ್ಬ ಘ್ಬ ಙ್ಬ - ಚ್ಬ ಛ್ಬ ಜ್ಬ ಝ್ಬ ಞ್ಬ - ಟ್ಬ ಠ್ಬ ಡ್ಬ ಢ್ಬ ಣ್ಬ
ತ್ಬ ಥ್ಬ ದ್ಬ ಧ್ಬ ನ್ಬ - ಪ್ಬ ಫ್ಬ ಬ್ಬ ಭ್ಬ ಮ್ಬ - ಯ್ಬ ರ್ಬ ಲ್ಬ ವ್ಬ ಳ್ಬ

ma: X + ◌್ VIRAMA + ಮ mā:

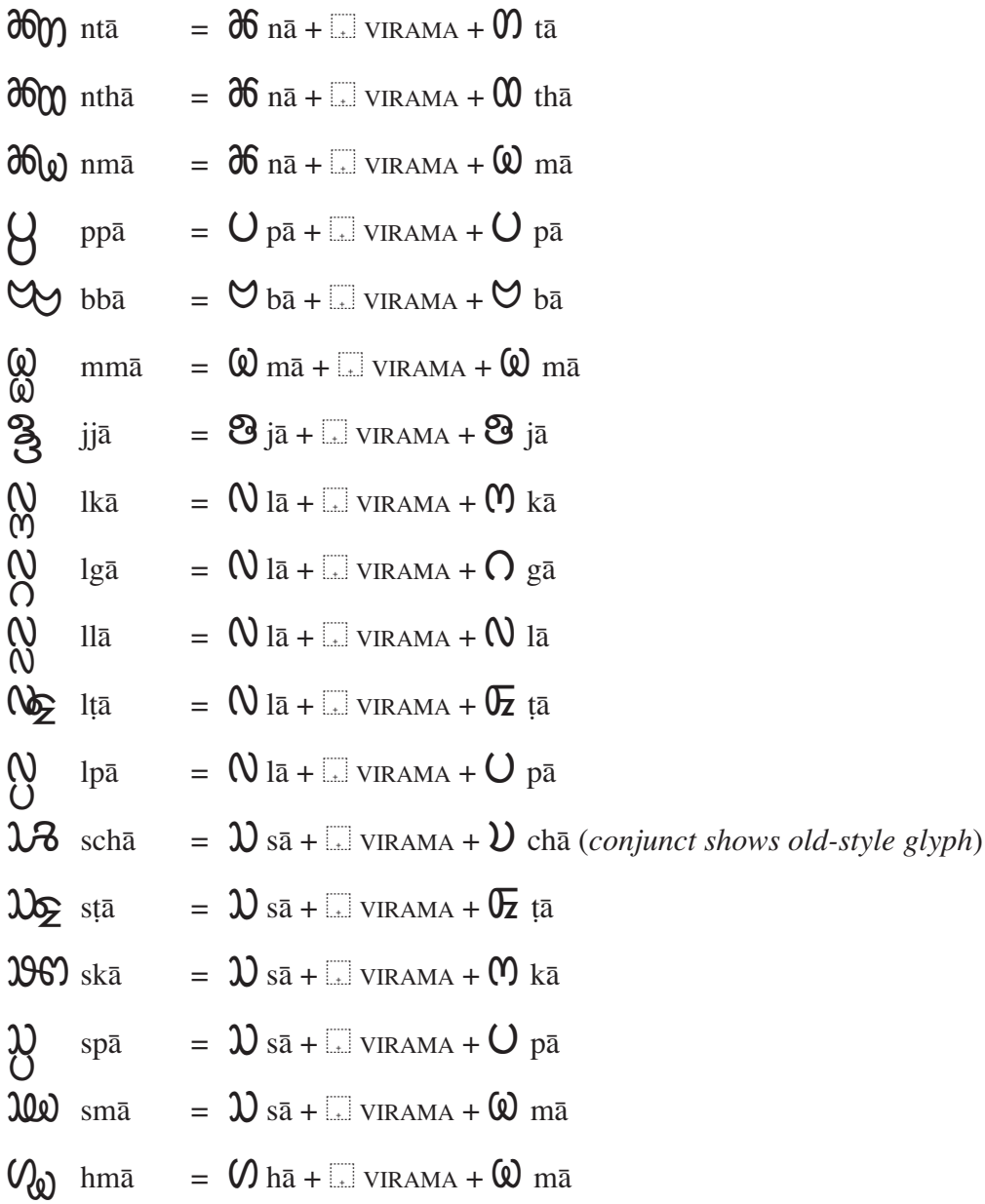
ಕ್ಮ ಖ್ಮ ಗ್ಮ ಘ್ಮ ಙ್ಮ - ಚ್ಮ ಛ್ಮ ಜ್ಮ ಝ್ಮ ಞ್ಮ - ಟ್ಮ ಠ್ಮ ಡ್ಮ ಢ್ಮ ಣ್ಮ
ತ್ಮ ಥ್ಮ ದ್ಮ ಧ್ಮ ನ್ಮ - ಪ್ಮ ಫ್ಮ ಬ್ಮ ಭ್ಮ ಮ್ಮ - ಯ್ಮ ರ್ಮ ಲ್ಮ ವ್ಮ ಳ್ಮ

ha: X + ◌್ VIRAMA + ಹ hā:

ಕ್ಹ ಖ್ಹ ಗ್ಹ ಘ್ಹ ಙ್ಹ - ಚ್ಹ ಛ್ಹ ಜ್ಹ ಝ್ಹ ಞ್ಹ - ಟ್ಹ ಠ್ಹ ಡ್ಹ ಢ್ಹ ಣ್ಹ
ತ್ಹ ಥ್ಹ ದ್ಹ ಧ್ಹ ನ್ಹ - ಪ್ಹ ಫ್ಹ ಬ್ಹ ಭ್ಹ ಮ್ಹ - ಯ್ಹ ರ್ಹ ಲ್ಹ ವ್ಹ ಳ್ಹ

In the 1982 book *Cāṅmār āg pudhi* a much wider range of conjunct pairs is shown, some of them with fairly complicated glyphs.

ಕ್ಕ kkā = ಕ kā + ◌್ VIRAMA + ಕ kā
ಕ್ಟ kṭā = ಕ kā + ◌್ VIRAMA + ಟ ṭā
ಕ್ತ ktā = ಕ kā + ◌್ VIRAMA + ತ tā
ಕ್ಮ kmā = ಕ kā + ◌್ VIRAMA + ಮ mā
ಕ್ಚ kcā = ಕ kā + ◌್ VIRAMA + ಚ cā (*conjunct shows old-style glyph*)
ಙ್ಕ ṅkā = ಙ ṅā + ◌್ VIRAMA + ಕ kā
ಙ್ಗ ṅgā = ಙ ṅā + ◌್ VIRAMA + ಗ gā
ಚ್ಚ ccā = ಚ cā + ◌್ VIRAMA + ಚ cā (*conjunct shows old-style glyph*)
ಚ್ಛ cchā = ಚ cā + ◌್ VIRAMA + ಛ chā (*conjunct shows old-style glyph*)
ಞ್ಚ ñcā = ಞ ñā + ◌್ VIRAMA + ಚ cā (*conjunct shows old-style glyph*)
ಞ್ಜ ñjā = ಞ ñā + ◌್ VIRAMA + ಜ jā
ಞ್ಝ ñjhā = ಞ ñā + ◌್ VIRAMA + ಝ jhā
ಟ್ಟ ṭṭā = ಟ ṭā + ◌್ VIRAMA + ಟ ṭā
ತ್ತ ttā = ತ tā + ◌್ VIRAMA + ತ tā
ತ್ಮ tmā = ತ tā + ◌್ VIRAMA + ಮ mā
ತ್ಥ tthā = ತ tā + ◌್ VIRAMA + ಥ thā
ಡ್ಡ ddā = ಡ dā + ◌್ VIRAMA + ಡ dā
ಡ್ಢ ddhā = ಡ dā + ◌್ VIRAMA + ಢ dhā



5.1 Specific recommendation for Chakma fonts. In Chakma, the encoding model supports conjunct behaviour and Chakma fonts *by default* should display the subjoined form of letters when following virama, to ensure legibility. Whether a conjunct is required or not is part of the spelling of a word; it is not a stylistic issue. (We have seen no examples of conjuncts with more than one consonant, and while the encoding handles (in principle) any length of stacking examples of such would probably be spelling errors.)

6. Collating order. As an Indo-European language, the standard Brahmic sorting order applies to Chakma.

7. Character names. Consonant letter names use the typical Brahmic transliteration used in the UCS. Chakma letters have a descriptive name followed by a traditional Brahmic consonant. These latter are given in annotations to the character names.

8. Punctuation and digits. Alongside a । DANDA and ॥ DOUBLE DANDA punctuation, Chakma has a unique ؞ QUESTION MARK, and a ꠰ SECTION MARK. There is some variation in the glyphs for the SECTION MARK, some looking like flowers or leaves. A set of digits exists and is encoded, although Bengali digits are also used. The Tanchangya use Myanmar digits.

9. Linebreaking. Letters and digits behave as in Bengali. Both CHAKMA DANDA and CHAKMA DOUBLE DANDA behave as in Devanagari. The CHAKMA QUESTION MARK behaves like U+003F QUESTION MARK. The CHAKMA SECTION MARK behaves like U+2055 FLOWER PUNCTUATION MARK.

10. Unicode Character Properties.

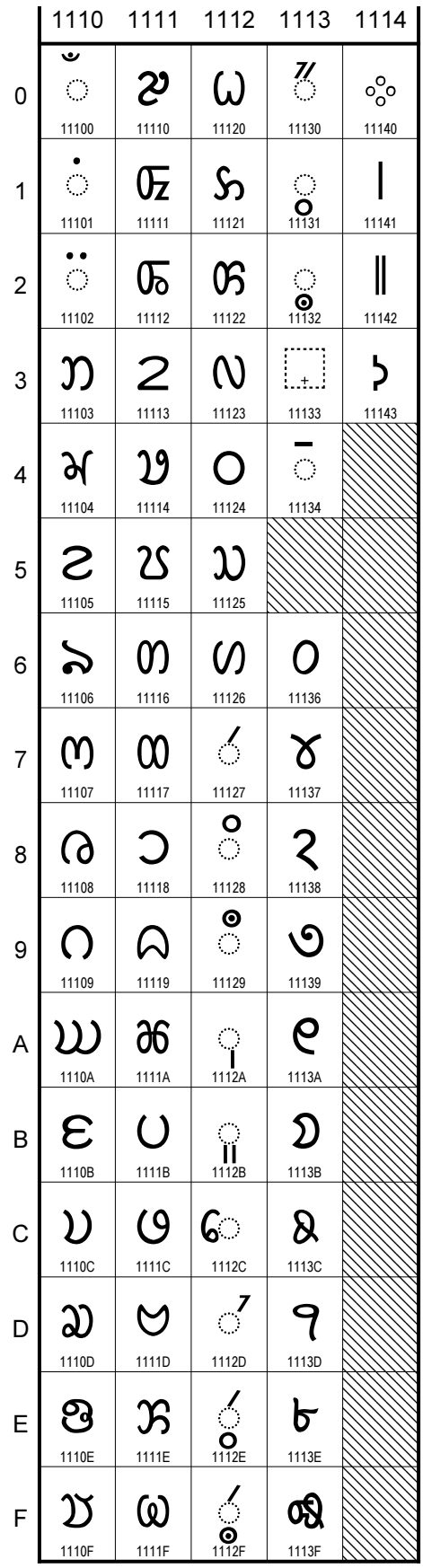
```
11100;CHAKMA SIGN CANDRABINDU;Mn;230;NSM;;;;;N;;;;;
11101;CHAKMA SIGN ANUSVARA;Mn;230;NSM;;;;;N;;;;;
11102;CHAKMA SIGN VISARGA;Mc;230;L;;;;;N;;;;;
11103;CHAKMA LETTER AA;Lo;0;L;;;;;N;;;;;
11104;CHAKMA LETTER I;Lo;0;L;;;;;N;;;;;
11105;CHAKMA LETTER U;Lo;0;L;;;;;N;;;;;
11106;CHAKMA LETTER E;Lo;0;L;;;;;N;;;;;
11107;CHAKMA LETTER KAA;Lo;0;L;;;;;N;;;;;
11108;CHAKMA LETTER KHAA;Lo;0;L;;;;;N;;;;;
11109;CHAKMA LETTER GAA;Lo;0;L;;;;;N;;;;;
1110A;CHAKMA LETTER GHAA;Lo;0;L;;;;;N;;;;;
1110B;CHAKMA LETTER NGAA;Lo;0;L;;;;;N;;;;;
1110C;CHAKMA LETTER CAA;Lo;0;L;;;;;N;;;;;
1110D;CHAKMA LETTER CHAA;Lo;0;L;;;;;N;;;;;
1110E;CHAKMA LETTER JAA;Lo;0;L;;;;;N;;;;;
1110F;CHAKMA LETTER JHAA;Lo;0;L;;;;;N;;;;;
11110;CHAKMA LETTER NYAA;Lo;0;L;;;;;N;;;;;
11111;CHAKMA LETTER TTAA;Lo;0;L;;;;;N;;;;;
11112;CHAKMA LETTER TTHAA;Lo;0;L;;;;;N;;;;;
11113;CHAKMA LETTER DDAA;Lo;0;L;;;;;N;;;;;
11114;CHAKMA LETTER DDHAA;Lo;0;L;;;;;N;;;;;
11115;CHAKMA LETTER NNAA;Lo;0;L;;;;;N;;;;;
11116;CHAKMA LETTER TAA;Lo;0;L;;;;;N;;;;;
11117;CHAKMA LETTER THAA;Lo;0;L;;;;;N;;;;;
11118;CHAKMA LETTER DAA;Lo;0;L;;;;;N;;;;;
11119;CHAKMA LETTER DHAA;Lo;0;L;;;;;N;;;;;
1111A;CHAKMA LETTER NAA;Lo;0;L;;;;;N;;;;;
1111B;CHAKMA LETTER PAA;Lo;0;L;;;;;N;;;;;
1111C;CHAKMA LETTER PHAA;Lo;0;L;;;;;N;;;;;
1111D;CHAKMA LETTER BAA;Lo;0;L;;;;;N;;;;;
1111E;CHAKMA LETTER BHAA;Lo;0;L;;;;;N;;;;;
1111F;CHAKMA LETTER MAA;Lo;0;L;;;;;N;;;;;
11120;CHAKMA LETTER YYAA;Lo;0;L;;;;;N;;;;;
11121;CHAKMA LETTER YAA;Lo;0;L;;;;;N;;;;;
11122;CHAKMA LETTER RAA;Lo;0;L;;;;;N;;;;;
11123;CHAKMA LETTER LAA;Lo;0;L;;;;;N;;;;;
11124;CHAKMA LETTER WAA;Lo;0;L;;;;;N;;;;;
11125;CHAKMA LETTER SAA;Lo;0;L;;;;;N;;;;;
11126;CHAKMA LETTER HAA;Lo;0;L;;;;;N;;;;;
11127;CHAKMA VOWEL SIGN A;Mn;230;NSM;;;;;N;;;;;
11128;CHAKMA VOWEL SIGN I;Mn;230;NSM;;;;;N;;;;;
11129;CHAKMA VOWEL SIGN II;Mn;230;NSM;;;;;N;;;;;
1112A;CHAKMA VOWEL SIGN U;Mn;220;NSM;;;;;N;;;;;
1112B;CHAKMA VOWEL SIGN UU;Mn;220;NSM;;;;;N;;;;;
1112C;CHAKMA VOWEL SIGN E;Mc;224;L;;;;;N;;;;;
1112D;CHAKMA VOWEL SIGN AI;Mn;230;NSM;;;;;N;;;;;
1112E;CHAKMA VOWEL SIGN O;Mn;0;NSM;11131 11127;;;;N;;;;;
1112F;CHAKMA VOWEL SIGN AU;Mn;0;NSM;11132 11127;;;;N;;;;;
11130;CHAKMA VOWEL SIGN OI;Mn;230;NSM;;;;;N;;;;;
11131;CHAKMA O MARK;Mn;220;NSM;;;;;N;;;;;
11132;CHAKMA AU MARK;Mn;220;NSM;;;;;N;;;;;
11133;CHAKMA VIRAMA;Mn;9;NSM;;;;;N;;;;;
11134;CHAKMA MAAYYAA;Mn;9;NSM;;;;;N;;;;;
11136;CHAKMA DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;
11137;CHAKMA DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
11138;CHAKMA DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
11139;CHAKMA DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
1113A;CHAKMA DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
1113B;CHAKMA DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
1113C;CHAKMA DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
1113D;CHAKMA DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
1113E;CHAKMA DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
1113F;CHAKMA DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;
11140;CHAKMA SECTION MARK;Po;0;L;;;;;N;;;;;
11141;CHAKMA DANDA;Po;0;L;;;;;N;;;;;
11142;CHAKMA DOUBLE DANDA;Po;0;L;;;;;N;;;;;
11143;CHAKMA QUESTION MARK;Po;0;L;;;;;N;;;;;
```

11. Bibliography.

Cāṅmā, Cirajyoti and Maṅgal Cāṅgmā. 1982. *Cāṅmār āg pudhi* (*Chakma primer*). Rāṅamāṭi: Cāṅmābhāṣā Prakāśanā Pariṣad.

Khisa, Bhagadatta. 2001. *Cāṅmā pattham pāt = Chakma primer*. Rāṅamāṭi: Tribal Cultural Institute (TCI).
Singā. 2004. *Phagadāṅ*.

12. Acknowledgements. This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at UC Berkeley) in respect of the Chakma encoding. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.



	1110	1111	1112	1113	1114
0	11100	11110	11120	11130	11140
1	11101	11111	11121	11131	11141
2	11102	11112	11122	11132	11142
3	11103	11113	11123	11133	11143
4	11104	11114	11124	11134	
5	11105	11115	11125		
6	11106	11116	11126	11136	
7	11107	11117	11127	11137	
8	11108	11118	11128	11138	
9	11109	11119	11129	11139	
A	1110A	1111A	1112A	1113A	
B	1110B	1111B	1112B	1113B	
C	1110C	1111C	1112C	1113C	
D	1110D	1111D	1112D	1113D	
E	1110E	1111E	1112E	1113E	
F	1110F	1111F	1112F	1113F	

Various signs

11100 ◌̐ CHAKMA SIGN CANDRABINDU
= caanaphupudaa
11101 ◌̇ CHAKMA SIGN ANUSVARA
= ekaphudaa
11102 ◌̈ CHAKMA SIGN VISARGA
= dviphudaa

Independent vowels

11103 𑄃 CHAKMA LETTER AA
= pichapujhaa aa
11104 𑄄 CHAKMA LETTER I
= delabhaangagaa i
11105 𑄅 CHAKMA LETTER U
= bacacu u
11106 𑄆 CHAKMA LETTER E
= lejaubaa e

Consonants

11107 𑄇 CHAKMA LETTER KAA
= cucyaangyaa kaa
11108 𑄈 CHAKMA LETTER KHAA
= grajaangyaa khaa
11109 𑄉 CHAKMA LETTER GAA
= caandyaa gaa
1110A 𑄊 CHAKMA LETTER GHAA
= tinaddaalyaa ghaa
1110B 𑄋 CHAKMA LETTER NGAA
= cilaama ngaa
1110C 𑄌 CHAKMA LETTER CAA
= dvibhalyaa caa
1110D 𑄍 CHAKMA LETTER CHAA
= majaraa chaa
1110E 𑄎 CHAKMA LETTER JAA
= dvipadalaa haa
1110F 𑄏 CHAKMA LETTER JHAA
= uraauraa jhaa
11110 𑄐 CHAKMA LETTER NYAA
= silaacyaa nyaa
11111 𑄑 CHAKMA LETTER TTAA
= dviyaadaat ttaa
11112 𑄒 CHAKMA LETTER TTHAA
= phudaadviyaat tthaa
11113 𑄓 CHAKMA LETTER DDAA
= aadudaangaat ddaa
11114 𑄔 CHAKMA LETTER DDHAA
= lejabharaat ddhaa
11115 𑄕 CHAKMA LETTER NNAA
= pettttuyaa nnaa
11116 𑄖 CHAKMA LETTER TAA
= ghangadaat taa
11117 𑄗 CHAKMA LETTER THAA
= jagadaat thaa
11118 𑄘 CHAKMA LETTER DAA
= dolaniit daa
11119 𑄙 CHAKMA LETTER DHAA
= talamuyaat dhaa
1111A 𑄚 CHAKMA LETTER NAA
= phaarabaanyaa naa
1111B 𑄛 CHAKMA LETTER PAA
= paalyaa paa
1111C 𑄜 CHAKMA LETTER PHAA
= ubaraphudaa phaa
1111D 𑄝 CHAKMA LETTER BAA
= ubaramuyaa baa
1111E 𑄞 CHAKMA LETTER BHAA
= ciraddaalyaa bhaa
1111F 𑄟 CHAKMA LETTER MAA
= bugatpadalaa maa
11120 𑄠 CHAKMA LETTER YYAA
= cimayyaa yyaa
11121 𑄡 CHAKMA LETTER YAA
= jilyaa yaa
11122 𑄢 CHAKMA LETTER RAA
= dvidaayyaa raa
11123 𑄣 CHAKMA LETTER LAA
= talamuyaa laa
11124 𑄤 CHAKMA LETTER WAA
= bajhonyaa waa
11125 𑄥 CHAKMA LETTER SAA
= bhudibukyaa saa
11126 𑄦 CHAKMA LETTER HAA
= ubaramuyaa haa

Dependent vowel signs

11127 ◌𑄀 CHAKMA VOWEL SIGN A
= ubaratulyaa a
11128 ◌𑄁 CHAKMA VOWEL SIGN I
= bahryaa i
11129 ◌𑄂 CHAKMA VOWEL SIGN II
= baaniiphadaa ii
1112A ◌𑄃 CHAKMA VOWEL SIGN U
= ekattaana u
1112B ◌𑄄 CHAKMA VOWEL SIGN UU
= dvittaana uu
1112C ◌𑄅 CHAKMA VOWEL SIGN E
= ekaara e
1112D ◌𑄆 CHAKMA VOWEL SIGN AI
= delabhaanga ai
1112E ◌𑄇 CHAKMA VOWEL SIGN O
= okaara o
1112F ◌𑄈 CHAKMA VOWEL SIGN AU
= aukaara au
11130 ◌𑄉 CHAKMA VOWEL SIGN OI
= oikaara oi
11131 ◌𑄊 CHAKMA O MARK
11132 ◌𑄋 CHAKMA AU MARK

Various signs

11133 ◌𑄍 CHAKMA VIRAMA
• used to form conjuncts
→ 1039 ◌္ myanmar sign virama
11134 ◌𑄎 CHAKMA MAAYYAA
• killer
→ 103A ◌် myanmar sign asat

Digits

11136 𑄶 CHAKMA DIGIT ZERO
11137 𑄷 CHAKMA DIGIT ONE
11138 𑄸 CHAKMA DIGIT TWO
11139 𑄹 CHAKMA DIGIT THREE
1113A 𑄺 CHAKMA DIGIT FOUR
1113B 𑄻 CHAKMA DIGIT FIVE
1113C 𑄼 CHAKMA DIGIT SIX
1113D 𑄽 CHAKMA DIGIT SEVEN
1113E 𑄾 CHAKMA DIGIT EIGHT

1113F 𑄯 CHAKMA DIGIT NINE

Punctuation

11140 𑄀 CHAKMA SECTION MARK
= phulacihna
11141 । CHAKMA DANDA
= ekacilyaa
11142 ॥ CHAKMA DOUBLE DANDA
= dvicilyaa
11143 𑄁 CHAKMA QUESTION MARK
= pujhaar

DR. G. A. Grierson সংগৃহীত চাকমা বর্ণ(১৯০৩ খৃ:)এবং এগুলি সম্বন্ধে তাঁর মতামত

The following account of the Chākmā alphabet is based on information provided by Dewan Kristo Chandra, a gentleman of Chākmā nationality, and forwarded to me by Mr. J. A. Cave-Browne, Assistant Commissioner, Chittagong Hill Tracts.

The Chākmā alphabet is as follows:—

Bengali.

323 BENGALI.

The most important point to notice in this alphabet is that the vowel inherent in each consonant is, not *a* as in other Indian languages, but *ā*. Note also that 𑄃 the initial form (there is, of course, no non-initial form) of *ā* is treated as a consonant, much as the letter *alif* is treated as a consonant in Arabic.

For purposes of comparison, I here give the usual Burmese forms of the consonants:—

က ka, ခ kha, ဂ ga, ဃ gha, င ṅa,
စ cha, ဆ chha, ဇ ja, ၎ jha, ဉ ña,
ဋ ṭa, ဌ ṭha, ဍ ḍa, ဎ ḍha, ဏ ṇa,
တ ta, ထ tha, ဒ da, ဓ dha, န na,
ပ pa, ဖ pha, ဗ ba, ဘ bha, မ ma,
ယ ya, ရ ra, လ la, ဝ wa
သ sa, ဟ ha

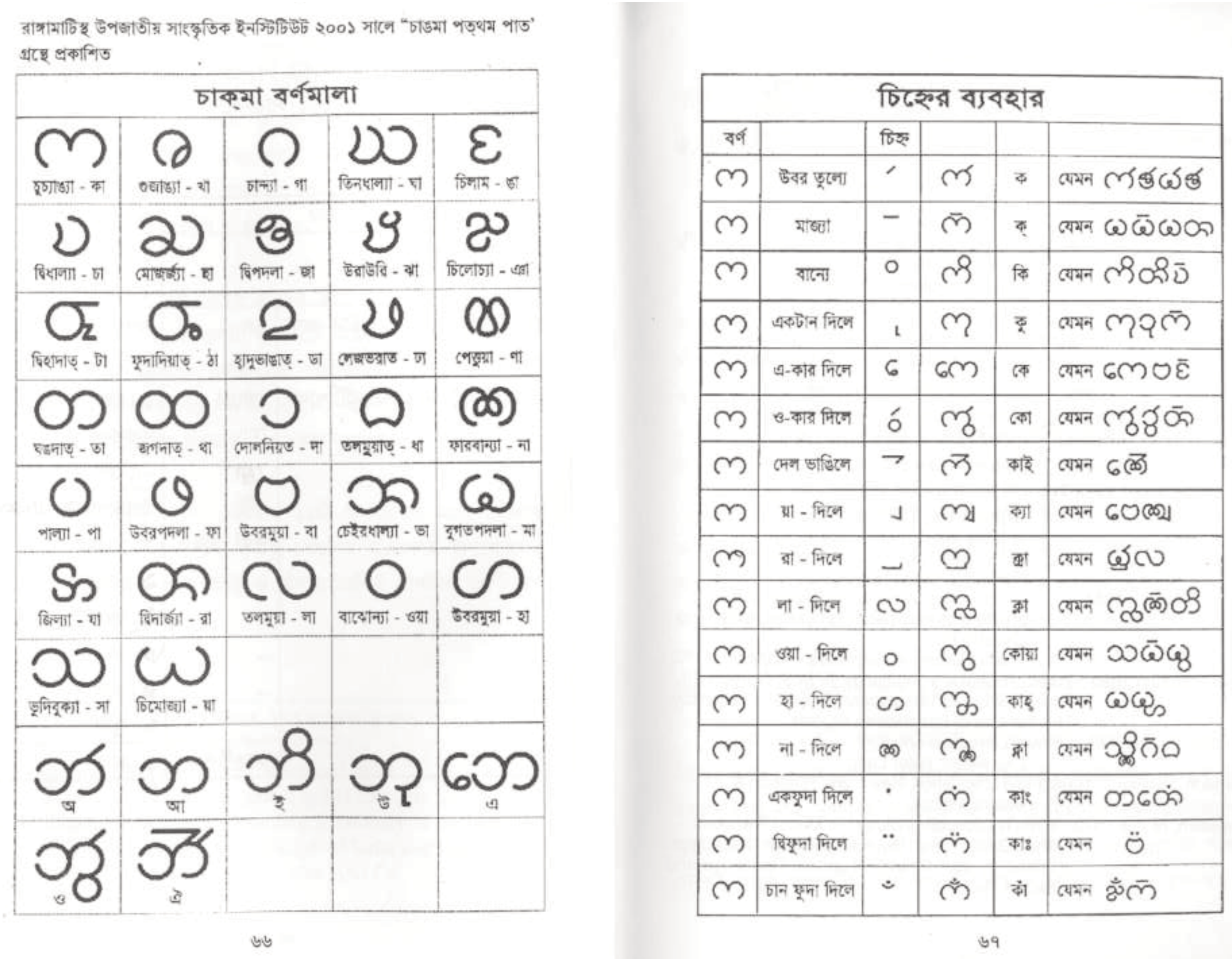


Figure 2. Charts taken from a paper written by Mr Sugata Chakma of the Tribal Cultural Institute, on "the Primary classification of languages".

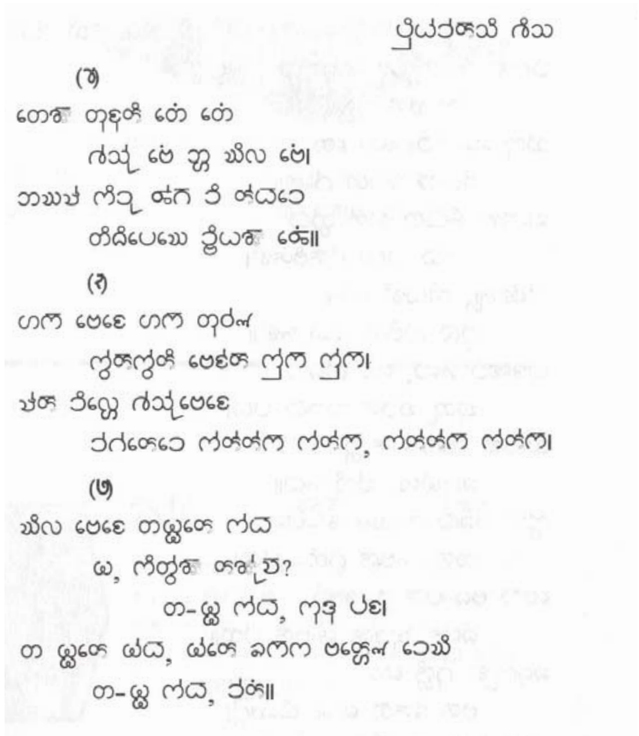


Figure 3. Example of poetry from the book *Phagadāṅ*, 2004.

Figure 4. Alphabet chart from Khisa 2001.

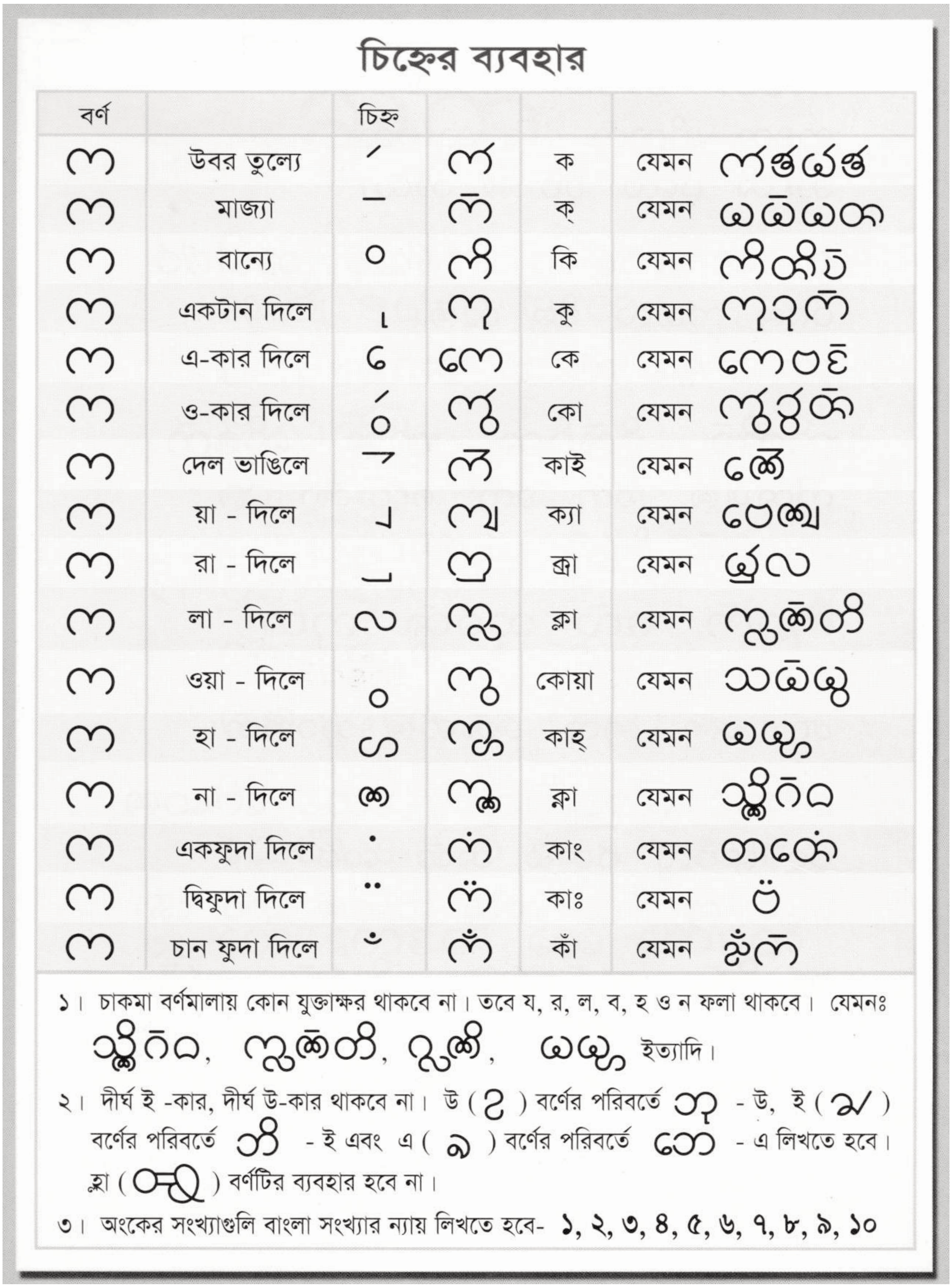


Figure 5. Chart of vowel signs and conjuncts from Khisa 2001.

চাঙমার আগ্ পুধি

(জদা অক্ষ্যর)

Figure 6. Chart with old-style conjuncts from Cāṅmā 1982.

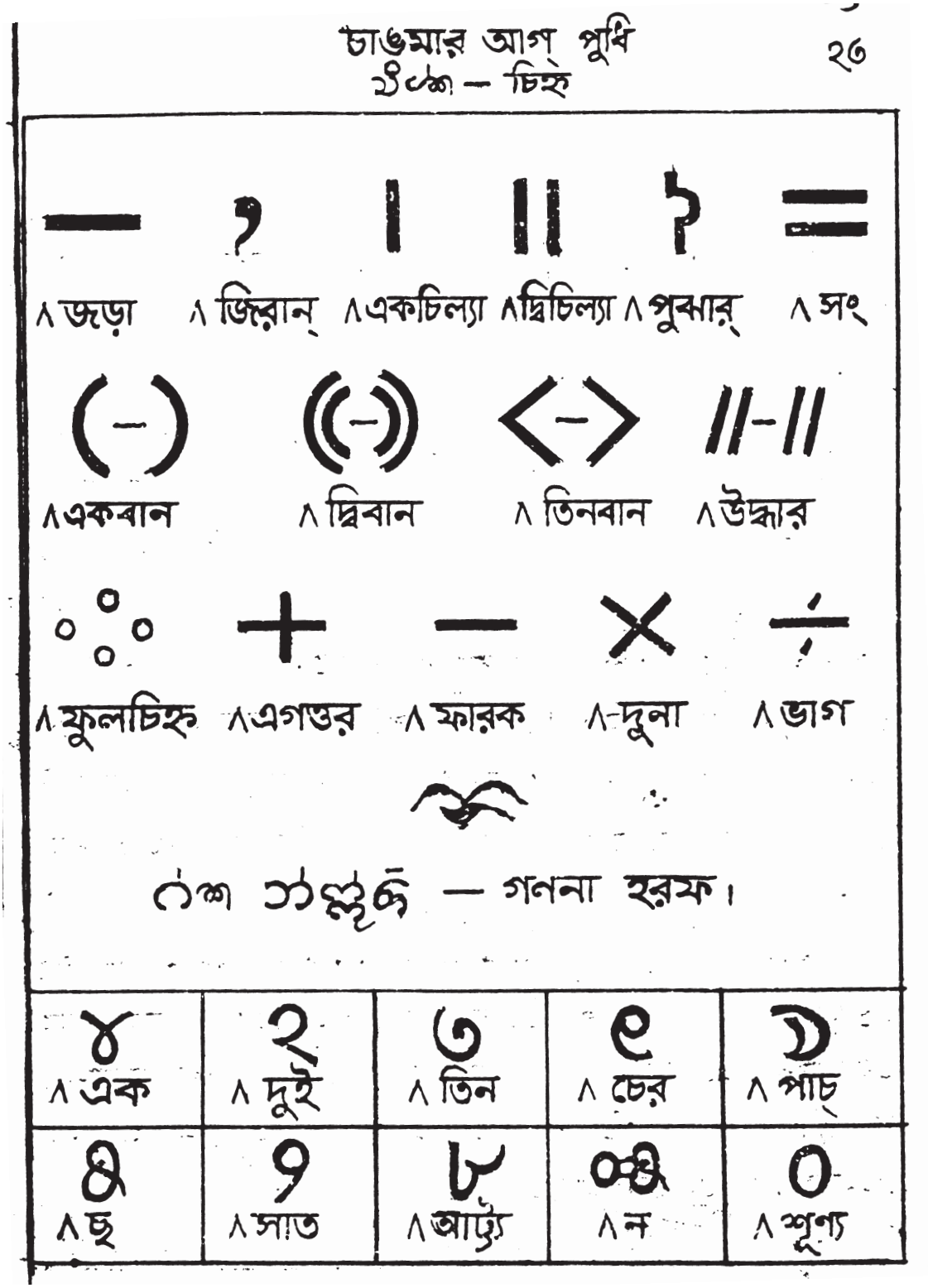


Figure 7. Chart with punctuation and digits from Cāṅmā 1982. Shown are: — EM DASH, , COMMA, । CHAKMA DANDA, ॥ CHAKMA DOUBLE DANDA, ؟ CHAKMA QUESTION MARK, = EQUALS SIGN; (LEFT PARENTHESIS,) RIGHT PARENTHESIS, ((LEFT DOUBLE PARENTHESIS,)) RIGHT DOUBLE PARENTHESIS, < LESS-THAN SIGN, > GREATER-THAN SIGN (unless these are single guillemets ‹ › or angle brackets 〈 〉), SOLIDUS repeated (the name given is *udkār* 'quotation' so these could likely be quotation marks “”); ꠶ CHAKMA SECTION MARK, + PLUS SIGN, − MINUS SIGN, × MULTIPLICATION SIGN, and ÷ DIVISION SIGN.

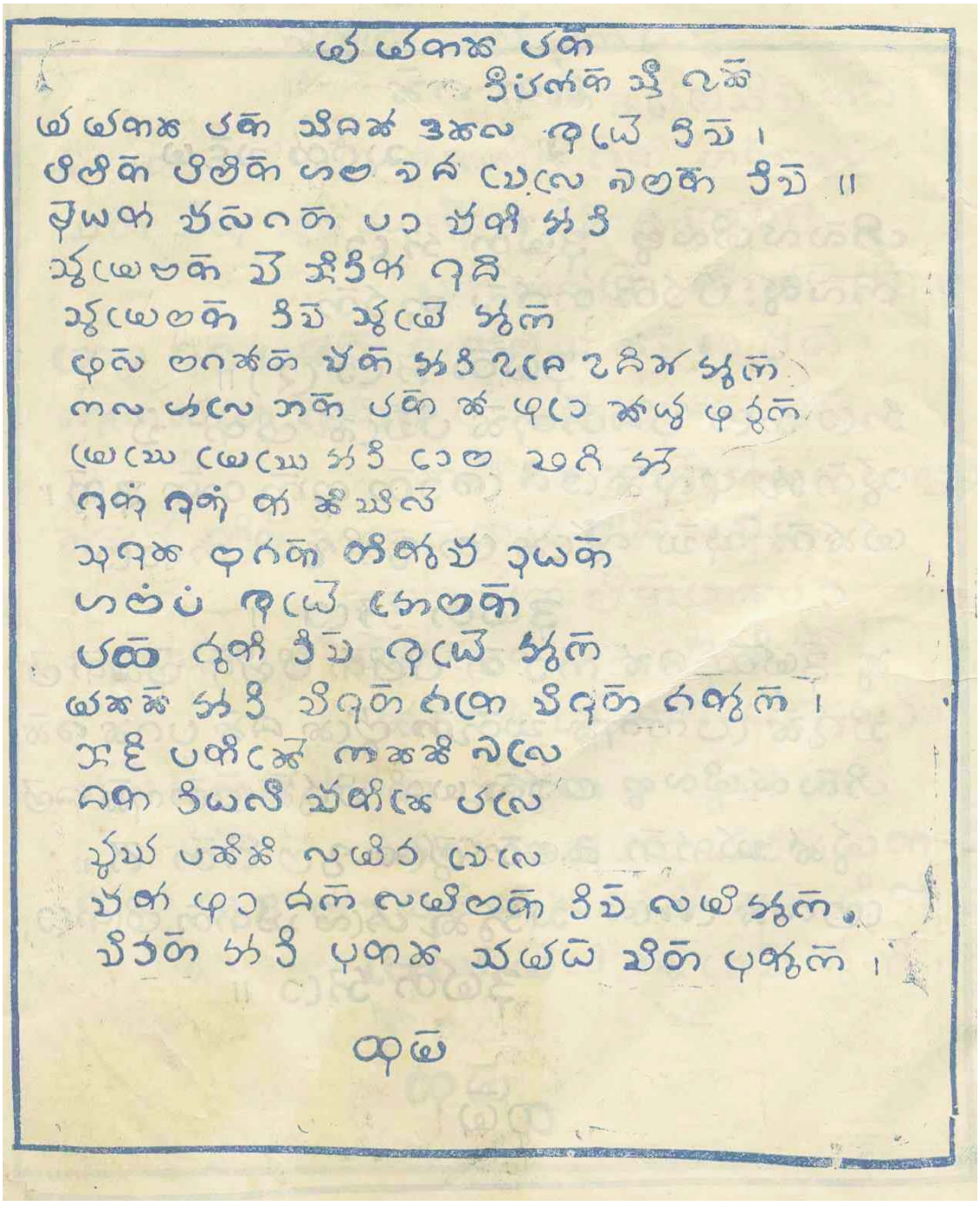


Figure 8. Handwritten text with DANDA and DOUBLE DANDA.

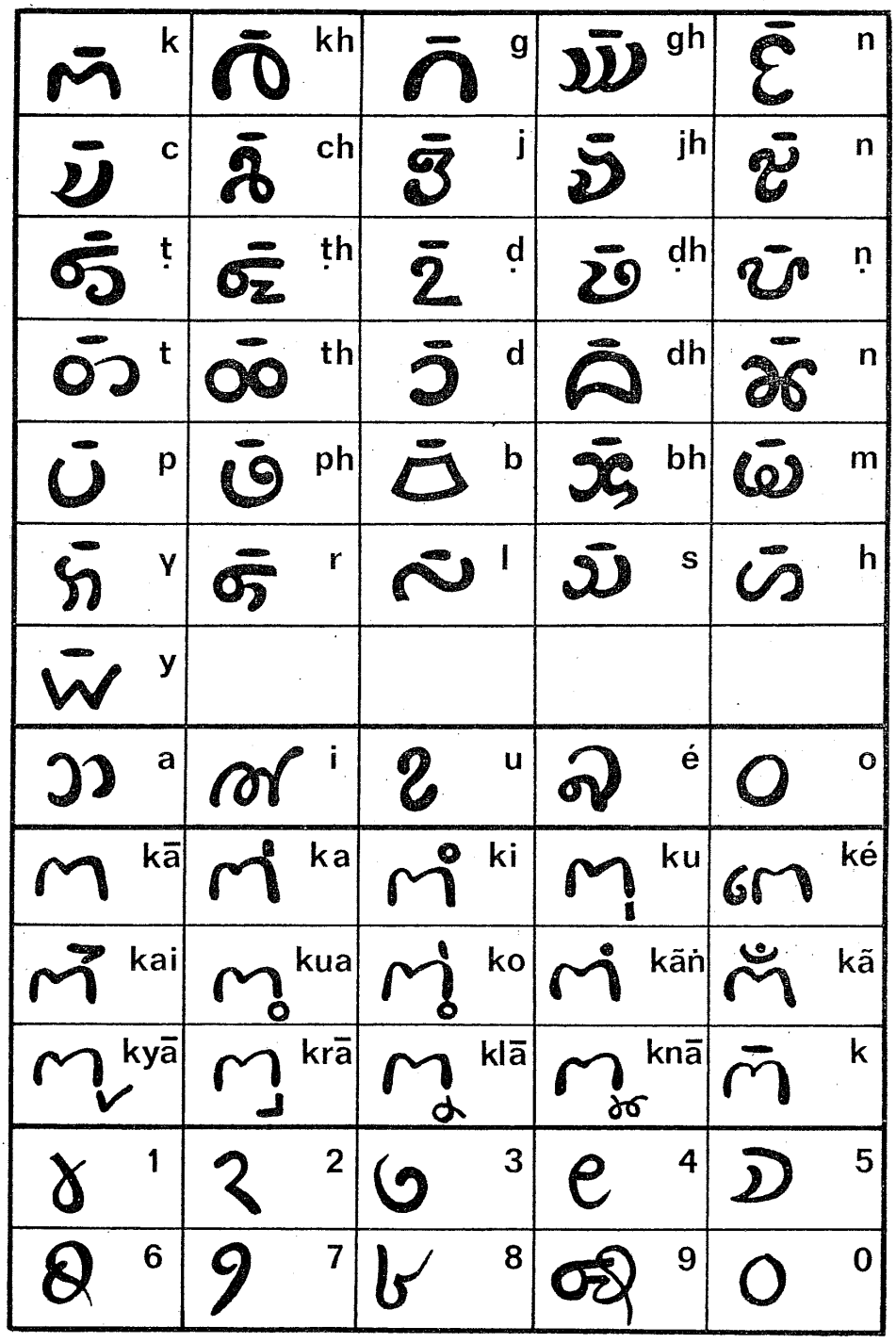


Figure 9. Chakma alphabet chart from Bernot 1972.

A. Administrative

1. Title

Proposal for encoding the Chakma script in the UCS

2. Requester's name

UC Berkeley Script Encoding Initiative (Universal Scripts Project)

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

Liaison contribution.

4. Submission date

2009-07-28

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

No.

6b. More information will be provided later

Yes.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

Yes.

1b. Proposed name of script

Chakma.

1c. The proposal is for addition of character(s) to an existing block

No.

1d. Name of the existing block

2. Number of characters in proposal

67.

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

Category A.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson and Hangendra Chakma.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes.

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?

Hagendra Chakma, Provungshu Chakma, John Clifton, Keisuke Huziwara, Pragya Joyoti, Saikat Khisa, Helen Leake, Chandra Roy

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?
People living in Bangladesh and in India.
4a. The context of use for the proposed characters (type of use; common or rare)
Common.
4b. Reference
5a. Are the proposed characters in current use by the user community?
Yes.
5b. If YES, where?
In Bangladesh and in India.
6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?
Yes.
6b. If YES, is a rationale provided?
Yes.
6c. If YES, reference
Contemporary use and accordance with the Roadmap.
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?
Yes.
8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?
No.
8b. If YES, is a rationale for its inclusion provided?
8c. If YES, reference
9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?
No.
9b. If YES, is a rationale for its inclusion provided?
9c. If YES, reference
10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?
No.
10b. If YES, is a rationale for its inclusion provided?
10c. If YES, reference
11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?
No.
11b. If YES, is a rationale for such use provided?
11c. If YES, reference
11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
No.
11e. If YES, reference
12a. Does the proposal contain characters with any special properties such as control function or similar semantics?
No.
12b. If YES, describe in detail (include attachment if necessary)
13a. Does the proposal contain any Ideographic compatibility character(s)?
No.
13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?