

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 Lepcha script in the BMP of the UCS****Source: Michael Everson****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2005-07-04**

Lepcha is a Sino-Tibetan language. The Lepchas call themselves *Róngkup* ('children of the Róng'), and their language *Róngring* ('language of the Róng'), but the name *Lepcha* is very widespread now, and the Lepchas themselves also use it frequently. The Lepcha script is in current use by many people in Sikkim and West Bengal, especially in the Darjeeling district. The weekly government newspaper *The Sikkim Herald* is also published in Lepcha and Lepcha is taught in schools in Sikkim. In West Bengal, there are many evening classes in which the script is taught, as well as books and magazines which are published privately or through one of several Lepcha associations. The SIL Ethnologue indicates that there are 41,300 Lepchas.

Structure

The Lepcha script is of the Brahmic type, but indirectly, being derived directly from Tibetan; some of those innovations are reflected in the structure of Lepcha. If one turns a Lepcha syllable clockwise, the similarities to Tibetan syllables is often quite striking. Lepcha has also introduced some innovations of its own, such as the diacritical marks used for final consonants. Lepcha is thought to have been devised around 1720 CE by the Sikkim king Phyag-rdor rNam-rgyal ("Chakdor Namgyal", born 1686), according to Jensen 1969.

Consonants bear the inherent vowel, but no virama is used to kill this vowel; vowel matras modify it, and explicit final consonants are used where there is no inherent vowel. Initial vowels are represented with the vowel matras on the neutral letter ࠄ A. Initial consonants can be followed by the glides ࠄ -YA and ࠄ -RA, both of which normally ligate with the consonant they modify; these can also combine to form ࠄ -rya, which is simply a glyph ligature of the other two: ࠄ kya, ࠄ kra, ࠄ krya. The glide -la is also found, but is represented not by a ligating combining mark, but by a limited set of letters containing this glide inherently. With few exceptions, these "combined" letters do not look like a ligature of their base letters with some mark: ࠄ ka ࠄ kla, ࠄ ga ࠄ gla, ࠄ pa ࠄ pla, ࠄ fa ࠄ fla, ࠄ ba ࠄ bla, ࠄ ma ࠄ mla, ࠄ ha ࠄ hla.

Three retroflex consonants exist in Lepcha. The syllables ࠄ KRA, ࠄ HRA, and ࠄ GRA are used to express these, though it is common to use the NUKTA to distinguish them from ordinary [kra], [hra], and [gra]: ࠄ KRA [ṭa], ࠄ HRA [ṭha], and ࠄ GRA [ḍa]. In Mainwaring's grammar, he also suggested that the dot be used with ࠄ ZA and ࠄ RA— ࠄ RA (*dra*), ࠄ ZA (*dza*)— but this does not seem to have been taken up by Lepcha writers. Chakraborty gives a different set of nuktated letters making use of ࠄ KRA [ṭa] and ࠄ GRA [ḍa], but with ࠄ HRA for [ḍha] and ࠄ KHRA for [ṭha]. Recently however, works have been published in Lepcha using three new letters: ࠄ TTA [ṭa], ࠄ TTHA [ṭha], and ࠄ DDA [ḍa].

A syllable is structured (and represented in the backing store) as C(·)(R)(Y)(V)(^)(F); the initial consonant “C” may be nuktated (·), and may be followed by -RA (R) or -YA (Y) or both (as *-rya*, never as **-yra*), an optional vowel matra (V), the diacritical mark RAN (^)—only after the inherent vowel or *-a* or *-i*, never with any of the other vowels—and optionally a final consonant (F). Of the finals, two of them represent the velar nasal *-ñ*; 𑄧 NYIN-DO is used only when there is no vowel matra; 𑄨 KANG is used in all other instances. (In the section on ordering below, the syllables with NYIN-DO are underlined>. The two are encoded separately because as named characters they may be required in isolation in discussions of the writing system; positional variants would not support this efficiently.

Glyph placement

The vowel signs in Lepcha follow the consonant to which they are applied. Three of them are rendered before the consonant: 𑄧 *ki*, 𑄨 *ko*, 𑄩 *kō*; one of them is drawn below the consonant: 𑄪 *ke*. When accompanying 𑄫 -A or 𑄬 -I, the modifier RAN rests atop the base consonant, though it follows them in the backing store: 𑄭 *KĀ*. In those syllables with RAN which also have final consonants above the base consonant, the RAN is rendered above the final: 𑄮 *KĀM*. Where the final consonant KANG co-occurs with 𑄫 -I, 𑄬 -O, or 𑄭 -OO, it is rendered to the left of them, though it follows them in the backing store: 𑄯 KANG, 𑄰 KING. The backing store for 𑄱 KRYĪNG is 𑄨 KA + 𑄫 -RA + 𑄬 -YA + 𑄭 -I + 𑄨 RAN + 𑄨 KANG.

Digits and punctuation

Digits have distinctive forms. Nowadays, the Lepchas use traditional punctuation signs only when copying the old books, but in everyday writing, they use European marks like COMMA, FULL STOP, and QUESTION MARK, though sometimes Tibetan TSHEGS are found. (Lepcha fonts should have a glyph for U+0F0D for this occasional usage.) The traditional punctuation marks are also encoded here. The names given here are found in Mainwaring & Grünwedel 1979, though they are not linked to glyphs there: 𑄲 TA-ROL ‘bar, crossbar of a loom, stop (in writing)’, 𑄳 NYET THYOOM TA-ROL ‘double bar’, 𑄴 CER-WA ‘cross-mark’, 𑄵 TSHOOK CER-WA ‘stop cross-mark’, 𑄶 TSHOOK ‘stop’ (*cf.* Tibetan *tsheg*). The Lepcha TA-ROLS are usually considerably shorter than Devanagari dandas, and as Lepcha text is very often found alongside Devanagari text, a unification would be inappropriate. The Lepcha TA-ROLS also have other variant shapes – such as having dots below them – which are unknown in Devanagari.

Implementations

Ka’ōnohi Kai implemented Lepcha for Mac OS X 10.2 in 2003, based on a font by Jason Glavy. This commercial product was based on my exploratory Unicode Lepcha proposal (published at www.evertype.com/standards/tai/lepcha.pdf). The encoding model and repertoire used by XenoType is a subset of the model used in *this* proposal; it is clear that the model works well for Lepcha.

Collating order

A syllabic ordering applies to Lepcha. Each initial consonant is given with its vowel and with all the consonant finals, then the same consonant is given with the next vowel and with all the consonant finals, and so on. After -E the the glides -YA and -RA are added to the base consonant and the vowels and consonant finals are run through again. The end of a vowel run is marked below by a semicolon; the transcribed syllable of the beginning of each vowel run is given in **bold type**.

𑄨 KA, 𑄨 KAK, 𑄨 KAM, 𑄨 KAL, 𑄨 KAN, 𑄨 KAP, 𑄨 KAR, 𑄨 KAT, 𑄯 KANG; 𑄨 KĀ 𑄨 KĀK, 𑄨 KĀM, 𑄨 KĀL, 𑄨 KĀN, 𑄨 KĀP, 𑄨 KĀR, 𑄨 KĀT, 𑄨 KĀNG; 𑄨 KÁ, 𑄨 KÁK, 𑄨 KÁM, 𑄨 KÁL, 𑄨 KÁN, 𑄨 KÁP, 𑄨 KÁR, 𑄨 KÁT, 𑄨 KÁNG; 𑄨 KĀ 𑄨 KĀK, 𑄨 KĀM, 𑄨 KĀL, 𑄨 KĀN, 𑄨 KĀP, 𑄨 KĀR, 𑄨 KĀT, 𑄨 KĀNG; 𑄧 KI,

ʃĕ KIK, ʃĕ KIM, ʃĕ KIL, ʃĕ KIN, ʃĕ KIP, ʃĕ KIR, ʃĕ KIT, ʃĕ KING; ʃĕ KĪ, ʃĕ KĪK, ʃĕ KĪM, ʃĕ KĪL,
 ʃĕ KĪN, ʃĕ KĪP, ʃĕ KĪR, ʃĕ KĪT, ʃĕ KĪNG; ʃĕ KO, ʃĕ KOK, ʃĕ KOM, ʃĕ KOL, ʃĕ KON, ʃĕ KOP, ʃĕ KOR,
 ʃĕ KOT, ʃĕ KONG; ʃĕ KÓ, ʃĕ KÓK, ʃĕ KÓM, ʃĕ KÓL, ʃĕ KÓN, ʃĕ KÓP, ʃĕ KÓR, ʃĕ KÓT, ʃĕ KÓNG;
 ʃĕ KU, ʃĕ KUK, ʃĕ KUM, ʃĕ KUL, ʃĕ KUN, ʃĕ KUP, ʃĕ KUR, ʃĕ KUT, ʃĕ KUNG; ʃĕ KÚ, ʃĕ KÚK,
 ʃĕ KÚM, ʃĕ KÚL, ʃĕ KÚN, ʃĕ KÚP, ʃĕ KÚR, ʃĕ KÚT, ʃĕ KÚNG; ʃĕ KE, ʃĕ KEK, ʃĕ KEM, ʃĕ KEL,
 ʃĕ KEN, ʃĕ KEP, ʃĕ KER, ʃĕ KET, ʃĕ KENG; ʃĕ KYA, ʃĕ KYAK, ʃĕ KYAM, ʃĕ KYAL, ʃĕ KYAN, ʃĕ KYAP,
 ʃĕ KYAR, ʃĕ KYAT, ʃĕ KYANG; ʃĕ KYÁ, ʃĕ KYÁK, ʃĕ KYÁM, ʃĕ KYÁL, ʃĕ KYÁN, ʃĕ KYÁP, ʃĕ KYÁR,
 ʃĕ KYÁT, ʃĕ KYÁNG; ʃĕ KYÁ, ʃĕ KYÁK, ʃĕ KYÁM, ʃĕ KYÁL, ʃĕ KYÁN, ʃĕ KYÁP, ʃĕ KYÁR,
 ʃĕ KYÁT, ʃĕ KYÁNG; ʃĕ KYÁ, ʃĕ KYÁK, ʃĕ KYÁM, ʃĕ KYÁL, ʃĕ KYÁN, ʃĕ KYÁP, ʃĕ KYÁR,
 ʃĕ KYÁT, ʃĕ KYÁNG; ʃĕ KYI, ʃĕ KYIK, ʃĕ KYIM, ʃĕ KYIL, ʃĕ KYIN, ʃĕ KYIP, ʃĕ KYIR,
 ʃĕ KYIT, ʃĕ KYING; ʃĕ KYÍ, ʃĕ KYÍK, ʃĕ KYÍM, ʃĕ KYÍL, ʃĕ KYÍN, ʃĕ KYÍP, ʃĕ KYÍR,
 ʃĕ KYÍT, ʃĕ KYÍNG; ʃĕ KYO, ʃĕ KYOK, ʃĕ KYOM, ʃĕ KYOL, ʃĕ KYON, ʃĕ KYOP, ʃĕ KYOR,
 ʃĕ KYOT, ʃĕ KYONG; ʃĕ KYÓ, ʃĕ KYÓK, ʃĕ KYÓM, ʃĕ KYÓL, ʃĕ KYÓN, ʃĕ KYÓP, ʃĕ KYÓR,
 ʃĕ KYÓT, ʃĕ KYÓNG; ʃĕ KYU, ʃĕ KYUK, ʃĕ KYUM, ʃĕ KYUL, ʃĕ KYUN, ʃĕ KYUP, ʃĕ KYUR,
 ʃĕ KYUT, ʃĕ KYUNG; ʃĕ KYÚ, ʃĕ KYÚK, ʃĕ KYÚM, ʃĕ KYÚL, ʃĕ KYÚN, ʃĕ KYÚP, ʃĕ KYÚR,
 ʃĕ KYÚT, ʃĕ KYÚNG; ʃĕ KYE, ʃĕ KYEK, ʃĕ KYEM, ʃĕ KYEL, ʃĕ KYEN, ʃĕ KYEP, ʃĕ KYER, ʃĕ KYET,
 ʃĕ KYENG; ʃĕ KRA, ʃĕ KRAK, ʃĕ KRAM, ʃĕ KRAL, ʃĕ KRAN, ʃĕ KRAP, ʃĕ KRAR, ʃĕ KRAT, ʃĕ KRANG;
 ʃĕ KRÁ, ʃĕ KRÁK, ʃĕ KRÁM, ʃĕ KRÁL, ʃĕ KRÁN, ʃĕ KRÁP, ʃĕ KRÁR, ʃĕ KRÁT, ʃĕ KRÁNG; ʃĕ KRÁ,
 ʃĕ KRÁK, ʃĕ KRÁM, ʃĕ KRÁL, ʃĕ KRÁN, ʃĕ KRÁP, ʃĕ KRÁR, ʃĕ KRÁT, ʃĕ KRÁNG; ʃĕ KRÁ,
 ʃĕ KRÁK, ʃĕ KRÁM, ʃĕ KRÁL, ʃĕ KRÁN, ʃĕ KRÁP, ʃĕ KRÁR, ʃĕ KRÁT, ʃĕ KRÁNG; ʃĕ KRI, ʃĕ KRIK, ʃĕ KRIM,
 ʃĕ KRIL, ʃĕ KRIN, ʃĕ KRIP, ʃĕ KRIR, ʃĕ KRIT, ʃĕ KRING; ʃĕ KRÍ, ʃĕ KRÍK, ʃĕ KRÍM, ʃĕ KRÍL,
 ʃĕ KRÍN, ʃĕ KRÍP, ʃĕ KRÍR, ʃĕ KRÍT, ʃĕ KRÍNG; ʃĕ KRO, ʃĕ KROK, ʃĕ KROM, ʃĕ KROL, ʃĕ KRON,
 ʃĕ KROP, ʃĕ KROR, ʃĕ KROT, ʃĕ KRONG; ʃĕ KRÓ, ʃĕ KRÓK, ʃĕ KRÓM, ʃĕ KRÓL, ʃĕ KRÓN, ʃĕ KRÓP,
 ʃĕ KRÓR, ʃĕ KRÓT, ʃĕ KRÓNG; ʃĕ KRU, ʃĕ KRUK, ʃĕ KRUM, ʃĕ KRUL, ʃĕ KRUN, ʃĕ KRUP,
 ʃĕ KRUR, ʃĕ KRUT, ʃĕ KRUNG; ʃĕ KRÚ, ʃĕ KRÚK, ʃĕ KRÚM, ʃĕ KRÚL, ʃĕ KRÚN, ʃĕ KRÚP,
 ʃĕ KRÚR, ʃĕ KRÚT, ʃĕ KRÚNG; ʃĕ KRE, ʃĕ KREK, ʃĕ KREM, ʃĕ KREL, ʃĕ KREN, ʃĕ KREP, ʃĕ KREK,
 ʃĕ KRET, ʃĕ KRENG; ʃĕ KRYA, ʃĕ KRYAK, ʃĕ KRYAM, ʃĕ KRYAL, ʃĕ KRYAN, ʃĕ KRYAP, ʃĕ KRYAR,
 ʃĕ KRYAT, ʃĕ KRYANG; ʃĕ KRYÁ, ʃĕ KRYÁK, ʃĕ KRYÁM, ʃĕ KRYÁL, ʃĕ KRYÁN, ʃĕ KRYÁP, ʃĕ KRYÁR,
 ʃĕ KRYÁT, ʃĕ KRYÁNG; ʃĕ KRYÁ, ʃĕ KRYÁK, ʃĕ KRYÁM, ʃĕ KRYÁL, ʃĕ KRYÁN, ʃĕ KRYÁP, ʃĕ KRYÁR,
 ʃĕ KRYÁT, ʃĕ KRYÁNG; ʃĕ KRYÁ, ʃĕ KRYÁK, ʃĕ KRYÁM, ʃĕ KRYÁL, ʃĕ KRYÁN, ʃĕ KRYÁP, ʃĕ KRYÁR,
 ʃĕ KRYÁT, ʃĕ KRYÁNG; ʃĕ KRYI, ʃĕ KRYIK, ʃĕ KRYIM, ʃĕ KRYIL,
 ʃĕ KRYIN, ʃĕ KRYIP, ʃĕ KRYIR, ʃĕ KRYIT, ʃĕ KRYING; ʃĕ KRYÍ, ʃĕ KRYÍK, ʃĕ KRYÍM, ʃĕ KRYÍL,
 ʃĕ KRYÍN, ʃĕ KRYÍP, ʃĕ KRYÍR, ʃĕ KRYÍT, ʃĕ KRYÍNG; ʃĕ KRYO, ʃĕ KRYOK,
 ʃĕ KRYOM, ʃĕ KRYOL, ʃĕ KRYON, ʃĕ KRYOP, ʃĕ KRYOR, ʃĕ KRYOT, ʃĕ KRYONG; ʃĕ KRYÓ,
 ʃĕ KRYÓK, ʃĕ KRYÓM, ʃĕ KRYÓL, ʃĕ KRYÓN, ʃĕ KRYÓP, ʃĕ KRYÓR, ʃĕ KRYÓT, ʃĕ KRYÓNG;

ཅྱ) KRYU, ཅྱ) KRYUK, ཅྱ) KRYUM, ཅྱ) KRYUL, ཅྱ) KRYUN, ཅྱ) KRYUP, ཅྱ) KRYUR, ཅྱ) KRYUT,
 ཅྱ) KRYUNG; ཅྱ) KRYÚ, ཅྱ) KRYÚK, ཅྱ) KRYÚM, ཅྱ) KRYÚL, ཅྱ) KRYÚN, ཅྱ) KRYÚP, ཅྱ) KRYÚR,
 ཅྱ) KRYÚT, ཅྱ) KRYÚNG; ཅྱ) KRYE, ཅྱ) KRYEK, ཅྱ) KRYEM, ཅྱ) KRYEL, ཅྱ) KRYEN, ཅྱ) KRYEP, ཅྱ) KRYER,
 ཅྱ) KRYET, ཅྱ) KRYENG ...

... and thus for KLA, for KHA, then GA, then NGA, and so on.

It would appear that the new series of retroflex consonants should be sorted in the customary position for these letters in Tibetan and other Brahmic scripts, so འ CA > ར CHA > ལ JA > ས NYA > ས TTA > ས THHA > ས DDA > ས TA > ས THA > ས DA > ས NA and so on. Whether these letters should appear in that position in the code table is uncertain; I would recommend not inserting them there pending ballot feedback.

Linebreaking

Opportunities for hyphenation occur after any full orthographic syllable. Lepcha punctuation marks can be expected to have behaviour similar to that of Devanagari DANDA and DOUBLE DANDA.

Unicode Character Properties

```

1C00;LEPCHA LETTER KA;Lo;0;L;;;;;N;;;;;
1C01;LEPCHA LETTER KLA;Lo;0;L;;;;;N;;;;;
1C02;LEPCHA LETTER KHA;Lo;0;L;;;;;N;;;;;
1C03;LEPCHA LETTER GA;Lo;0;L;;;;;N;;;;;
1C04;LEPCHA LETTER GLA;Lo;0;L;;;;;N;;;;;
1C05;LEPCHA LETTER NGA;Lo;0;L;;;;;N;;;;;
1C06;LEPCHA LETTER CA;Lo;0;L;;;;;N;;;;;
1C07;LEPCHA LETTER CHA;Lo;0;L;;;;;N;;;;;
1C08;LEPCHA LETTER JA;Lo;0;L;;;;;N;;;;;
1C09;LEPCHA LETTER NYA;Lo;0;L;;;;;N;;;;;
1C0A;LEPCHA LETTER TA;Lo;0;L;;;;;N;;;;;
1C0B;LEPCHA LETTER THA;Lo;0;L;;;;;N;;;;;
1C0C;LEPCHA LETTER DA;Lo;0;L;;;;;N;;;;;
1C0D;LEPCHA LETTER NA;Lo;0;L;;;;;N;;;;;
1C0E;LEPCHA LETTER PA;Lo;0;L;;;;;N;;;;;
1C0F;LEPCHA LETTER PLA;Lo;0;L;;;;;N;;;;;
1C10;LEPCHA LETTER PHA;Lo;0;L;;;;;N;;;;;
1C11;LEPCHA LETTER FA;Lo;0;L;;;;;N;;;;;
1C12;LEPCHA LETTER FLA;Lo;0;L;;;;;N;;;;;
1C13;LEPCHA LETTER BA;Lo;0;L;;;;;N;;;;;
1C14;LEPCHA LETTER BLA;Lo;0;L;;;;;N;;;;;
1C15;LEPCHA LETTER MA;Lo;0;L;;;;;N;;;;;
1C16;LEPCHA LETTER MLA;Lo;0;L;;;;;N;;;;;
1C17;LEPCHA LETTER TSA;Lo;0;L;;;;;N;;;;;
1C18;LEPCHA LETTER TSHA;Lo;0;L;;;;;N;;;;;
1C19;LEPCHA LETTER DZA;Lo;0;L;;;;;N;;;;;
1C1A;LEPCHA LETTER YA;Lo;0;L;;;;;N;;;;;
1C1B;LEPCHA LETTER RA;Lo;0;L;;;;;N;;;;;
1C1C;LEPCHA LETTER LA;Lo;0;L;;;;;N;;;;;
1C1D;LEPCHA LETTER HA;Lo;0;L;;;;;N;;;;;
1C1E;LEPCHA LETTER HLA;Lo;0;L;;;;;N;;;;;
1C1F;LEPCHA LETTER VA;Lo;0;L;;;;;N;;;;;
1C20;LEPCHA LETTER SA;Lo;0;L;;;;;N;;;;;
1C21;LEPCHA LETTER SHA;Lo;0;L;;;;;N;;;;;
1C22;LEPCHA LETTER WA;Lo;0;L;;;;;N;;;;;
1C23;LEPCHA LETTER A;Lo;0;L;;;;;N;;;;;
1C24;LEPCHA SUBJOINED LETTER YA;Mc;0;L;;;;;N;;;;;
1C25;LEPCHA SUBJOINED LETTER RA;Mc;0;L;;;;;N;;;;;
1C26;LEPCHA VOWEL SIGN AA;Mc;0;L;;;;;N;;;;;
1C27;LEPCHA VOWEL SIGN I;Mc;0;L;;;;;N;;;;;
1C28;LEPCHA VOWEL SIGN O;Mc;0;L;;;;;N;;;;;
1C29;LEPCHA VOWEL SIGN OO;Mc;0;L;;;;;N;;;;;
1C2A;LEPCHA VOWEL SIGN U;Mc;0;L;;;;;N;;;;;
1C2B;LEPCHA VOWEL SIGN UU;Mc;0;L;;;;;N;;;;;
1C2C;LEPCHA VOWEL SIGN E;Mn;0;NSM;;;;;N;;;;;
1C2D;LEPCHA CONSONANT SIGN K;Mn;0;NSM;;;;;N;;;;;
1C2E;LEPCHA CONSONANT SIGN M;Mn;0;NSM;;;;;N;;;;;
1C2F;LEPCHA CONSONANT SIGN L;Mn;0;NSM;;;;;N;;;;;
1C30;LEPCHA CONSONANT SIGN N;Mn;0;NSM;;;;;N;;;;;
1C31;LEPCHA CONSONANT SIGN P;Mn;0;NSM;;;;;N;;;;;
1C32;LEPCHA CONSONANT SIGN R;Mn;0;NSM;;;;;N;;;;;
1C33;LEPCHA CONSONANT SIGN T;Mn;0;NSM;;;;;N;;;;;
1C34;LEPCHA CONSONANT SIGN NYIN-DO;Mc;0;L;;;;;N;;;;;
1C35;LEPCHA CONSONANT SIGN KANG;Mc;0;L;;;;;N;;;;;
1C36;LEPCHA SIGN RAN;Mn;0;NSM;;;;;N;;;;;
1C37;LEPCHA SIGN NUKTA;Mn;7;NSM;;;;;N;;;;;
    
```

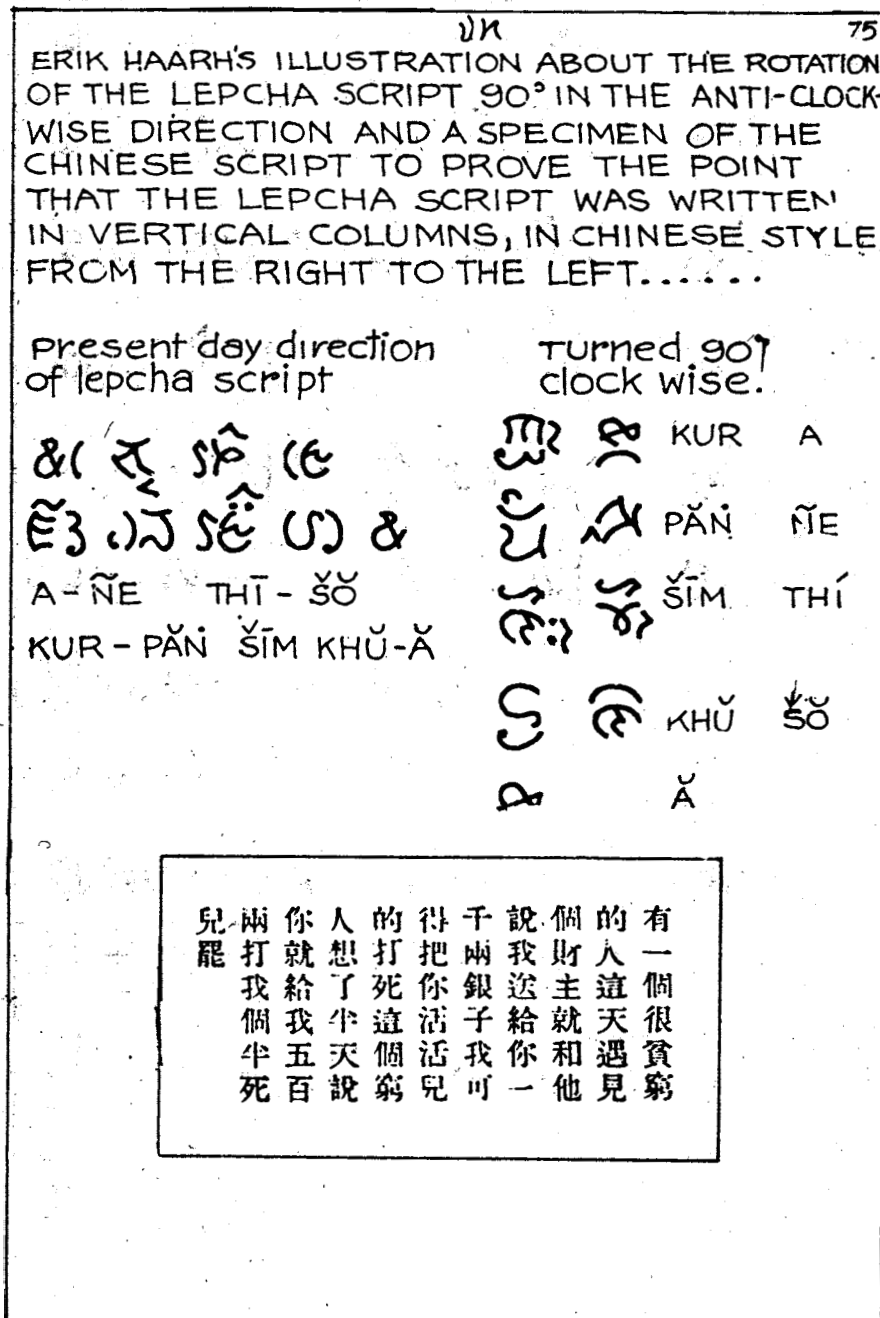



Figure 2. Discussion in Chakraborty 1978 about Lepcha letterforms. Apparently the *dbu-med* precursors of Lepcha were written in vertical columns, and subsequently when the lines were turned to left-to-right orientation, the letters took advantage of a new ductus to become Lepcha.



Figure 3. On the left, the usual modern glyph for ○(-AA). On the right, the glyph as found in older manuscripts. The manuscript -AA in particular is easier to distinguish from the glyph for (○ -O.

48

CHHOGYÚ KHÁ KÁ TSA LESSON
THIRTY NINE

KAKYÓT THAP

LETTERS LIKE ऽ; ॆ OR ॆ; ॆ; ॆ; ॆ HAVE BEEN USED IN LEPCHA TO REPRESENT CERTAIN OTHER FOREIGN SOUNDS. ORIGINALLY THEY WERE USED WITHOUT THE DOT. THE DOT WAS INTRODUCED BY MAIN WARING.

ॆ OR ॆ REPRESENT THE SOUND OR SIMPLY D ॆ. THUS ॆ OR ॆ IS PRONOUNCED AS DRÁN DRÓ OR DAN DÓ.

ॆ IS ALSO USED TO REPRESENT ॆ D OR ॆ WHICH IS NEAR TO TIBETAN GR (HARD SOUND). TIBETAN GR OR A HARD ROLLING TR OR SIMPLY ॆ IS REPRESENTED BY ॆ OR ॆ OR ॆ. THUS TIBETAN GRU (A SHIP) IS WRITTEN BY ॆ OR ॆ AND PRONOUNCED TRU OR TU.

ॆ IS PRONOUNCED AS DHRA OR DHA OR ॆ OR THRA OR THA OR ॆ. ॆ IS PRONOUNCED AS ॆ THRA, ॆ THA OR ॆ. ॆ IS PRONOUNCED AS ॆ THRA, ॆ THÁ OR ॆ (ACCORDING TO I MANSINGHA CHEMJONG)

READ :

ॆ	ॆ	ॆ ; ॆ ; ॆ	ॆ ; ॆ	ॆ ; ॆ	ॆ ; ॆ	ॆ	ॆ
ॆ	ॆ	ॆ ; ॆ	ॆ	ॆ	ॆ	ॆ	ॆ
TA	TA	TA ; THA	DA	DA	DHA	DHA	

READ AND TRANSCRIBE :

ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ)

ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ)

ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ) ॆ (ॆ ॆ ॆ)

Figure 4. Discussion in Chakraborty 1978 about using the NUKTA to represent retroflex sounds. Here ॆ HRA is used for [dha] and ॆ KHRA is used for [tha], though in most Lepcha writing ॆ HRA is used for [tha] and [dha] is not represented.

VOWELS IN INDO-TIBETAN SCRIPTS											
	COMMON ELEMENT		I	Í	U	Ú	E	AI/É	Ó	Ó	ANG
	A	Á									
TIBETAN	ཨ	ཨྱ	ཨི	ཨིྱ	ཨུ	ཨུྱ	ཨེ	ཨེ	ཨོ	ཨོྱ	ཨོྱ
TIBETAN UMEHIG	𑄎										
LIMBU OR KIRAT	𑄎	𑄎	𑄎			𑄎	𑄎, 𑄎	𑄎, 𑄎	𑄎, 𑄎	𑄎, 𑄎	𑄎
LEPCHA	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎		𑄎	𑄎/𑄎	𑄎
LEPCHA TURNED 90° CLOCK WISE	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎		𑄎	𑄎/𑄎	𑄎
BRAMHI	𑄎	𑄎	𑄎		𑄎		𑄎		𑄎		𑄎
PRO-LICH-CHAVI	𑄎	𑄎	𑄎	𑄎	𑄎		𑄎		𑄎	𑄎	𑄎
POST-LICH-CHAVI	𑄎	𑄎	𑄎	𑄎	𑄎		𑄎		𑄎	𑄎	𑄎
BENGALI	𑄎	𑄎	𑄎	𑄎	𑄎		𑄎		𑄎	𑄎	𑄎
KON KANI-DEVANAGARI	𑄎	𑄎	𑄎	𑄎	𑄎		𑄎		𑄎	𑄎	𑄎

K WITH DIFFERENT VOWELS											
	KA	KÁ	KI	KÍ	KU	KÚ	KE	KAI/KÉ	KO	KÓ	KANG
BRAMHI	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎
PRO-LICHCHAVI	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎
POST-LICHCHAVI	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎
TIBETAN	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎
LEPCHA	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎		𑄎	𑄎	𑄎
LEPCHA TURNED 90° CLOCK-WISE	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎		𑄎	𑄎	𑄎
BENGALI	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎	𑄎		𑄎	𑄎	𑄎

Figure 5. Chart from Chakraborty 1978 giving comparisons of Lepcha with other Brahmic scripts.

Lepcha		Tibetan		Power		Lepcha		Tibetan		Power		
printed type	written type	U-med	U-čan	Mainwaring	Our Transcription	printed type	written type	U-med	U-čan	Mainwaring	Our Transcription	
ཀ	ཀ	ཀ	ཀ	ka	kā	ཀ	ཀ			final	m	m
ཁ	ཁ	ཁ	ཁ	kla	klā	ཁ	ཁ			tsa	tsā	
			final	k	k	ཁ	ཁ			tsha	tšā	
ག	ག	ག	ག	kha	kā	ག	ག			za	zā	
གྷ	གྷ	གྷ	གྷ	ga	gā	གྷ	གྷ			ya	yā	
པ	པ			gla	glā	པ	པ			y	y	
ཅ	ཅ			nga	nā	ཅ	ཅ			ra	rā	
ཆ	ཆ		final	ng	n̄	ཆ	ཆ			final	r	r
ཇ	ཇ	ཇ	ཇ	cha	čā	ཇ	ཇ			r	r	
ཉ	ཉ	ཉ	ཉ	chha	čā	ཉ	ཉ			la	lā	
ཏ	ཏ	ཏ	ཏ	ja	jā	ཏ	ཏ			final	l	l
ཐ	ཐ	ཐ	ཐ	nya	nyā	ཐ	ཐ			ha	hā	
ད	ད	ད	ད	ta	tā	ད	ད			hla	hlā	
			final	t	t	ད	ད			va	vā	
དྷ	དྷ	དྷ	དྷ	tha	tā	དྷ	དྷ			sa	sā	
ན	ན	ན	ན	da	dā	ན	ན			sha	šā	
པ	པ	པ	པ	na	nā	པ	པ			wa	wā	
			final	n	n	པ	པ			a	ā	
ཕ	ཕ	ཕ	ཕ	pa	pā	ཕ	ཕ			á	á	
				pla	plā	ཕ	ཕ			á	á	
			final	p	p	ཕ	ཕ			i	i	
བ	བ	བ	བ	pha	pā	ཕ	ཕ			i	í	
				fa	fā	ཕ	ཕ			u	ü	
				fla	flā	ཕ	ཕ			ú	u	
མ	མ	མ	མ	ba	bā	ཕ	ཕ			e	e	
ཙ	ཙ	ཙ	ཙ	bla	blā	ཕ	ཕ			o	ó	
ཛ	ཛ	ཛ	ཛ	ma	mā	ཕ	ཕ			á	á	
ཞ	ཞ	ཞ	ཞ	m̄la	m̄lā	ཕ	ཕ					

Figure 6. Chart comparing printed and handwritten Lepcha with printed and *dbu-med* Tibetan letters, from Mainwaring & Grünwedel 1979 (1898).

Handwritten text in Lepcha script, consisting of approximately 12 lines of text within a rectangular border.

LEPCHA MANUSCRIPT, C. 1800 (INDIA OFFICE LIBRARY)
EXCERPTED FROM "ALPHABET - VOL. 2," P. 262. DAVID DIRINGER.

TENTATIVE RENDERING OF THE MANUSCRIPT PAGE INTO MODERN SCRIPT.

Handwritten text in Lepcha script, consisting of approximately 12 lines of text, presented as a tentative rendering of the manuscript page.

Figure 7. Manuscript c. 1800 written out in modern hand, from Chakraborty 1978. Chakraborty replaces the manuscript's TSHOOK punctuation with TA-ROL punctuation.

ཀུའི་ལྟོ་སྤྲི་ཏུ་ལྟོ་སྤྲི་ཏུ་

མཉམ་པས་བློ་གྲོས་བློ་གྲོས་ལྟོ་སྤྲི་ཏུ་ལྟོ་སྤྲི་ཏུ་
འཕྲུལ་ལྟོ་སྤྲི་ཏུ་ལྟོ་སྤྲི་ཏུ་
ལྟོ་སྤྲི་ཏུ་ལྟོ་སྤྲི་ཏུ་
ལྟོ་སྤྲི་ཏུ་ལྟོ་སྤྲི་ཏུ་

ལྟོ་སྤྲི་ཏུ་

Figure 10. Sample handwritten text from the literary periodical Aachuley, edited by Lyangsong Tamsang. Note how in this example, each of the TA-ROL punctuation marks bears a small dot beneath.

Table with 2 columns (20 and 29) showing Tibetan text and TA-ROL punctuation marks. A circled mark in column 20 has a dot underneath it.

Figure 11. Sample text from K. P. Tamsang’s Grammar of the Lepcha Language. Here true Tibetan TSHEGS are used.

1st.—THE CONSONANTS.

འཁོར་ ར་ ལ་ ས་ སྐ་ *amo re, 35.*

*K	Kh	G	Ng
ཅ	ཆ	ཇ	ཉ
Ch	Chh	J	Ny
ཏ	ཐ	ད	ལྷ
T	Th	D	N
ཎ	ཏ	*	ཏ
P	Ph	F	B
པ	ཕ	ཆ	ཇ
Ts	Tsh	Z	Y
ཅ	ཆ	ཇ	ཉ
R	L	H	V
ར	ལ	ཨ	ཐ
	S	Sh	W
	ས	ཨ	ཨ
Kl	Gl	Pl	Fl
ཏ	ཐ	ཕ	ཨ
	Bl	Ml	Hl
	བ	ཏ	ཨ

* All these consonants have the sound of the inherent short *a* affixed to them.

Figure 12. Chart of the Lepcha script from Mainwaring and Grünwedel 1979 (1898).
 Note the relative position of the consonants in series (ཅ YA, ར RA, ལ LA, ཨ HA, ཐ VA, ས SA, ཆ SHA, ཨ WA) with WA in last position.

DIACRITICAL MARKS,

called in Lepcha $\text{ཨྱ}(\text{ཨྱ})$ *thámbyn*, (implying the vowel and final signs, &c.).

$\text{ཨྱ}(\text{ཨྱ})$ ω $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ *ákup sa thámbyn kakyak gum,*

the Vowel Signs are seven in number, viz. :—

() () () () () () ()

These are united to ཨྱ a, the basis of all the vowels, as follows :

2nd.—THE EIGHT VOWELS.

* $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ *ákup kaku (8) re.*

			A			
			ཨྱ^\dagger			
Á	I	O	Ó	U	Ú	E
ཨྱ^\dagger	ཨྱ^\dagger	ཨྱ^\dagger	ཨྱ^\dagger	ཨྱ^\dagger	ཨྱ^\dagger	ཨྱ^\dagger

The vowel signs are similarly affixed to all the consonants.

3rd.—THE NINE FINALS.

$\text{ཨྱ}(\text{ཨྱ})$ ω $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ $\text{ཨྱ}(\text{ཨྱ})$ *Telbo sa thámbyn kakyót (9) re,*

the Final Signs are nine in number, and are thus formed,—

‘ “ ^ ~ o ~ - ‘)

* This name $\text{ཨྱ}(\text{ཨྱ})$ *ákup*, lit. child or small (letters), was formerly applied by the Lepchas exclusively to the Finals.

† This short a is inherent after all the consonants. The sound is effected by a simple (unaspirated) exhalation of the breath.

Figure 13. Chart of the Lepcha script from Mainwaring and Grünwedel 1979 (1898). Note the relative position of the vowels in series (A, Á, I, O, Ó, U, Ú, E).

CHHOGYU FALI THAP LESSON FOURTEEN

READ AND TRANSCRIBE:

Handwritten Lepcha syllabary chart for lesson 16, showing various combinations of consonants and vowels in a grid format.

Handwritten Lepcha syllabary chart for lesson 17, showing various combinations of consonants and vowels in a grid format.

READ AND TRANSCRIBE:

KA KA KÁ KÍ KO KÓ KU KÚ KE KANG
KHA KHÁ KHÍ KHÍ KHÓ KHÓ KHU KHÚ KHE KHANG
GA GÁ GÍ GÍ GO GÓ GU GÚ GE GANG
NGA NGÁ NGÍ NGÍ NGO NGÓ NU NGÚ NGE NANG
CHA CHÁ CHÍ CHÍ CHO CHÓ CHU CHÚ CHE CHANG
CHHA CHHÁ CHHÍ CHHÍ CHHO CHHÓ CHHU CHHÚ CHHE CHHANG
JA JÁ JÍ JÍ JO JÓ JU JÚ JE JANG
NYA NYÁ NYÍ NYÍ NYON NYÓ NYU NYÚ NYE NYANG

Figure 14. Chart from Chakraborty 1978 giving the Lepcha script in syllabary format. Note the relative position of the vowels in series (KA, KÁ, KÁ, KÍ, KÍ, KO, KÓ, KU, KÚ, KE, KANG) and the relative position of the consonants in series (WA, YA, RA, LA, HA, VA, SA, SHA) with WA unusually not in last position.

Vowels								
ᱠ	ᱡ	ᱢ	ᱣ	ᱤ	ᱥ	ᱦ	ᱧ	ᱨ
a	aa	ae	ai	ao	ao	au	ao	aey
Consonants								
ᱛ	ᱜ	ᱝ	ᱞ	ᱟ	ᱠ	ᱡ	ᱢ	ᱣ
ka	kha	ga	nga	ca	cha	ja	nya	
ᱤ	ᱥ	ᱦ	ᱧ	ᱨ	ᱩ	ᱪ	ᱫ	ᱬ
ta	tha	da	na	pa	pha	fa	ba	ma
ᱭ	ᱮ	ᱯ	ᱰ	ᱱ	ᱲ	ᱳ	ᱴ	ᱵ
tsha	tsha	za	ya	ra	la	ha	va	sa
ᱶ	ᱷ	ᱸ						
sya	sa	wa						

Figure 15. Chart of the Lepcha script from Tamsang 1978. Note the relative position of the vowels in series (A (a), Á (aa), I (ae), Í (ai), O (ao), Ó (ao), U (au), Ú (ao), E (aey)) and the relative position of the consonants in series (ᱛ YA, ᱜ RA, ᱝ LA, ᱞ HA, ᱟ VA, ᱠ SA, ᱡ SHA (sya), ᱢ WA) with WA in last position.

THE LEPCHA ALPHABET

TABLE 1: Consonants

ᱛ	ᱜ	ᱝ	ᱞ	ᱟ	ᱠ	ᱡ	ᱢ	ᱣ	
ka	kha	ga	nga	ca	cha	ja	nya		
ᱤ	ᱥ	ᱦ	ᱧ	ᱨ	ᱩ	ᱪ	ᱫ	ᱬ	
ta	tha	da	na	pa	pha	fa	ba	ma	
ᱭ	ᱮ	ᱯ	ᱰ	ᱱ	ᱲ	ᱳ	ᱴ	ᱵ	
tsha	tsha	za	ya	ra	la	ha	va		
ᱶ	ᱷ	ᱸ							
sha	sa	wa							
ᱹ	ᱺ	ᱻ	ᱼ	ᱽ	᱾	᱿	ᱽ		
kla	gla	pla	fla	bla	mha	hla			

TABLE 2: Syllabic vowels

ᱠ	ᱡ	ᱢ	ᱣ	ᱤ	ᱥ	ᱦ	ᱧ	ᱨ	ᱩ
a	á	â	i	í	o	ó	u	ú	e

TABLE 3: Diacritical vowel signs

ᱠ	ᱡ	ᱢ	ᱣ	ᱤ	ᱥ	ᱦ	ᱧ	ᱨ	ᱩ
a	á	â	i	í	o	ó	u	ú	e

Figure 16. Chart of the Lepcha script from Plaisier 2003. Note the relative position of the vowels in series (A, Á, Â, I, Í, O, Ó, U, Ú, E) and the relative position of the consonants in series (ᱛ YA, ᱜ RA, ᱝ LA, ᱞ HA, ᱟ VA, ᱠ SA, ᱡ SHA, ᱢ WA) with WA in last position.

TABLE XXX - Row 1C: LEPCHA

	1C0	1C1	1C2	1C3	1C4
0					
1					
2					
3					
4					
5					
6					
7					
8					
9					
A					
B					
C					
D					
E					
F					

G = 00
P = 00

A. Administrative

1. Title

Proposal for encoding the Lepcha script in the UCS.

2. Requester's name

Michael Everson

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

Individual contribution.

4. Submission date

2005-07-04

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

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

Yes.

Proposed name of script

Lepcha.

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

No.

1c. Name of the existing block

2. Number of characters in proposal

74.

3. Proposed category (see section II, Character Categories)

Category A.

4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

Level 2

4b. Is a rationale provided for the choice?

Yes.

4c. If YES, reference

Lepcha requires Level 2 implementation as other Brahmic scripts do.

5a. Is a repertoire including character names provided?

Yes.

5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

Yes.

5c. Are the character shapes attached in a legible form suitable for review?

Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Jason Glavy and Ka'ōnohi Kai via Michael Everson.

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

8. 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. See above.

9. 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.

Yes. See above.

C. Technical – Justification

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

Not to WG2. A preliminary proposal has been available at www.evertype.com/standards/tai/lepcha.pdf for a number of years. The UTC has seen a number of documents about Lepcha: L2/03-259 (Ka'ōnohi Kai, "Introduction to the Lepcha script"), L2/04-397 (Government of India, "Proposal to encode Lepcha"), and L2/05-061 (Michael Everson, "Analysis of the Indian proposal to encode Lepcha script in the UCS").

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?

Heleen Plaisier, Leiden University (linguist and script expert); Ka'ōnohi Kai, Xenotype Technologies (software implementor), Jason Glavy (font developer).

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?

Yes. See above.

4a. The context of use for the proposed characters (type of use; common or rare)

Common. Used to write the Lepcha language.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

Sikkim and West Bengal, and also in Bhutan and Nepal.

6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?

Yes. Positions 1C00-1C4F are proposed.

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)?

Yes.

11b. If YES, is a rationale for such use provided?

Yes.

11c. If YES, reference

Brahmic vowel signs.

12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

12b. If YES, reference

13a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

13b. If YES, describe in detail (include attachment if necessary)

14a. Does the proposal contain any Ideographic compatibility character(s)?

No.

14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?