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

Doc Type: Working Group Document === *DRAFT 2* ===
Title: Proposal to encode 10 Latin letters for North American indigenous orthographies
Source: Karl Pentzlin
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2010-06-28

To read this draft when distributed as Word document, the following fonts have to be installed:
"Roman Cyrillic Std" from http://kodeks.uni-bamberg.de/AKSL/Schrift/RomanCyrillicStd.htm
"Aboriginal Sans" and "Aboriginal Serif" from http://www.languagegeek.com/font/fontdownload.html

Additions for North American indigenous orthographies

χ   U+A794   LATIN CAPITAL LETTER CHI

          → 0058 latin capital letter x

χ   U+A795   LATIN SMALL LETTER CHI

          → 03C7 greek small letter chi
          · not intended for use with IPA

ʎ   U+A796   LATIN CAPITAL LETTER LAMBDA

          → 0245 latin capital letter turned v

λ   U+A797   LATIN SMALL LETTER LAMBDA

          · not intended for use with IPA

ƛ   U+A798   LATIN CAPITAL LETTER LAMBDA WITH STROKE

          · lower case is U+019B LATIN SMALL LETTER LAMBDA WITH STROKE

Ł   U+A799   LATIN CAPITAL LETTER L WITH BELT

          · lower case is U+026C LATIN SMALL LETTER L WITH BELT

Θ   U+A79A   LATIN CAPITAL LETTER THETA

          → 019F latin capital letter o with middle tilde
          → 03F4 greek capital theta symbol
          → 04E8 cyrillic capital letter barred o

θ   U+A79B   LATIN SMALL LETTER THETA

          → 0275 latin small letter barred o
          → 03B8 greek small letter theta
          · not intended for use with IPA

Ꝋ U+A79C   MODIFIER LETTER CAPITAL THETA
         → 1DBF modifier letter small theta

Addition for quality typographic representation of IPA

β U+A79D   LATIN SMALL LETTER BETA
         → 1DBF modifier letter small theta

Properties:

```
A794;LATIN CAPITAL LETTER CHI;Lu;0;L;;;;;N;;;;A795;
A795;LATIN SMALL LETTER CHI;Ll;0;L;;;;;N;;;A794;;A794
A796;LATIN CAPITAL LETTER LAMBDA;Lu;0;L;;;;;N;;;;A797;
A797;LATIN SMALL LETTER LAMBDA;Ll;0;L;;;;;N;;;A796;;A796
A798;LATIN CAPITAL LETTER LAMBDA WITH STROKE;Lu;0;L;;;;;N;;;;019B;
A799;LATIN CAPITAL LETTER L WITH BELT;Lu;0;L;;;;;N;;;;026C;
A79A;LATIN CAPITAL LETTER THETA;Lu;0;L;;;;;N;;;;A79B;
A79B;LATIN SMALL LETTER THETA;Ll;0;L;;;;;N;;;A79A;;A79A
A79C;MODIFIER LETTER CAPITAL THETA;Lm;0;L;<super> 03F4;;;;N;;;;;
A79D;LATIN SMALL LETTER BETA;Ll;0;L;;;;;N;;;;;
```

Properties to be changed for existing characters:

```
019B;LATIN SMALL LETTER LAMBDA WITH STROKE;Ll;0;L;;;;;N;LATIN SMALL LETTER BARRED LAMBDA;;A798;;A798
026C;LATIN SMALL LETTER L WITH BELT;Ll;0;L;;;;;N;LATIN SMALL LETTER L BELT;;A799;;A799
```

1. Introduction

Some letters, which were Greek in first line, are used now within Latin orthographies of living languages. In most cases the lowercase letters are used with slight modifications to adapt them to Latin typographic styles, and the uppercase letters were redesigned, usually as enlarged versions of the lowercase letter.

Thus, most of these are encoded separately as Latin letters.
This included all such letters used in African orthographies:

Alpha	α A:	Latin alpha	ɑ Ɑ	U+0251/2C6D
Gamma	γ Γ:	Latin gamma	ɣ Ɣ	U+0263/0194
Epsilon	ε E:	Latin open e	ɛ Ɛ	U+025B/0190
Iota	ɩ I:	Latin iota	ɪ ʟ	U+0269/0196
Upsilon	υ Y:	Latin upsilon	ʊ Ʊ	U+028A/01B1;
		Latin v with hook	ʋ Ʋ	U+028B/01B2

However, this is not the case for letters used in North American orthographies, which use (besides all the letters listed above, except alpha) letters based on theta, lambda, and chi. These letters apparently were introduced by their use in phonetics (Americanist and IPA). As such orthographies are accepted, the normal North American indigenous user does not care or know about the Greek origin, and wants to use it like any other letter of "his" script (i.e. the Latin script). While in the time of the introduction the typographical possibilities were limited for the normal user (using modified mechanical typewriters, at the best), this situation has be changed in these days where good typography is achievable on cheap computers for everyone.

Also, while some users did not use capital letters at all in the first time, this has changed now, especially when names and citations from indigenous languages are used in English text.

Therefore, there is a need to encode the letters originated from Greek as true Latin letters, together with their uppercase forms already developed, as it is already done for African languages.

While doing this, there also is a need to encode uppercase variant for other letters, which are introduced from phonetics (where they had no uppercase counterparts) but now are ordinary letters (for which uppercase forms are developed).

(The theta is also rumored to be used in some orthographies of Romani, a minority language used by gypsies in Europe, but this could not be confirmed at this time.)

Regarding the chi and the lambda, the uppercase form must be clearly distinguishable from:
    U+005F X LATIN CAPITAL LETTER X
    U+0245 Λ LATIN CAPITAL LETTER TURNED V.
The turned V is also used in North American orthographies.
Thus, their capital form must not use the uppercase form of the corresponding Greek letters. However, the tradition employed for the African letters, i.e. modeling the uppercase form after the lowercase form, is followed anyway.

This is also done for the theta, which yields a form more similar to U+03F4 ϴ GREEK CAPITAL THETA SYMBOL rather than U+0398 Θ GREEK CAPITAL LETTER THETA, which is in accordance to the fact that all other Latin capital letters do not have gaps between their parts.

The uppercase form of the modifier letter theta, which is use as a true letter in some orthographies, is modeled after the Latin letter proposed here.

Regarding the lambda, the names use the term LAMBDA according to the name used in English scientific and colloquial context, and resembling the part in the name of the already encoded Latin letter U+019B LATIN SMALL LETTER LAMBDA WITH STROKE, rather than the term LAMDA used in the Greek letters U+039B/03BB GREEK CAPITAL/SMALL LETTER LAMDA.

Having Latin chi, lambda, and theta, all Greek letters which gave the model for IPA letters are available as real Latin letters, except the beta. Therefore, a Latin beta is also included here. This makes IPA a real subset of the Latin script, allowing font creators to supply the complete IPA character set in a way compliant with high-quality Latin typography, while being able to include the complete Greek character set in the same font independently in a way compliant with high-quality Greek typography.

2. Usage of the proposed letters in North American indigenous languages

This list is based on a survey of http://www.languagegeek.com .

Theta:

Algonquin: 'Ɔ'ɔ́ɔ́ɔníííh (White Clay)

Coast Salish: Hǝṅq̓ǝmiṅǝm

Coast Salish: ʔayʔaǰuθǝm/Saɬuɬtxʷ (Sliammon Comox Homalco Klahoose)

Rotinonhsonni: Skarù·rę' (Tuscarora)

Rotinonhsonni: Wendat

Modifier Letter Theta:

Coast Salish: Nǝxʷsƛ̓ay̓ǝmúcǝn (Klallam)

Coast Salish: Ləк̓ʷiŋínəŋ (N Straits Salish)

Coast Salish: ʔayʔaǰuθəm/Saɬuɬtxʷ (Sliammon Comox Homalco Klahoose)

Chi:

Coast Salish: Hən̓q̓əmin̓əm

Coast Salish: Sliammon Comox Homalco Klahoose

Rotinonhsonni: Wendat

Lambda:

Wakashan: Haiɬzaqvla (Heiltsuk)

Lambda with stroke:

Coast Salish: Hən̓q̓əmin̓əm

Coast Salish: Nəxʷsƛ̓ay̓əmúcən (Klallam)

Coast Salish: Ləк̓ʷiŋínəŋ (N Straits Salish)

Coast Salish: ʔayʔaǰuθəm/Saɬuɬtxʷ (Sliammon Comox Homalco Klahoose)

Salish: Dəxʷləšucid

Salish: Nxaʔamxčín

Salish: Q̓ʷayáeɬq̓ (Upper Chehalis)

Wakashan: Haiɬzaqvla (Heiltsuk)

Wakashan: Qʷi·qʷi·diččaq (Makah)

Wakashan: Nuučaan̓uɬ

Rotinonhsonni: Wendat

L with belt:

Coast Salish: Hən̓q̓əmin̓əm

Coast Salish: Nəxʷsƛ̓ay̓əmúcən (Klallam)

Coast Salish: Sliammon Comox Homalco Klahoose

Turned V:

Iroquoian: Onʌyota'a:ka/Onʌyoteʔa·ka

3. Examples

Fig. 1: from http://www.languagegeek.com/salishan/sliammon.html , as of 2010-04-26
    showing the capital forms of "Theta" and "L with belt", and the Latin small form of "theta"

ʔAYʔAǰUΘƏM — SAɬUɬTXʷ

COMOX — HOMALCO — KLAHOOSE — SLIAMMON

ʔayʔaǰuθəm/Saɬuɬtxʷ Keyboards are available

Fig. 2: from *http://www.languagegeek.com/salishan/klallam.html , as of 2010-04-26*
showing the capital form of "lambda with stroke"

NƏXʷSƛAẎƏMÚCƏN
KLALLAM

Nəxʷsƛaẏəmúcən Keyboards are available

Fig. 3: from: Lepsius, C. R: Standard Alphabet for Reducing Unwritten Languages and Foreign Graphic Systems to a Uniform Orthography in European Letters. Second Edition, London/Berlin 1863.
The following excerpt from p.245 shows a similar form for a capital chi.

Specimen.

Χυαχο, χαχametwa, χ̓uči. Miṅkri χit̰irkin? Torekaite geal-
ḷm. Tawaχ warkin? Ketai, mintawaχom. Inan yarsnok hiwut
torekaite mịnrestịm. Meṅχostin? Doten χač. Neṅko χit̰irkin?
Inege χịlkit. Χυaχo iṅki! minχametwamịk kinemal.

Remarks.

The alphabet and the specimen are taken from the treatise
of L. Radloff: "*Ueber die Sprache der Tschuktschen*", in the
Mémoires de l'Acad. Impér. des Sciences. de St. Pétersbourg,
VIIᵉ série, tome III, nᵒ 10. 1861. The author of this treatise

Fig. 4: from: Handbook of the International Phonetic Association. Cambridge 1999, reprint 2003. ISBN 0 521 63751 1.
The cover page shown here shows a beta with a serif at the lower end, thus clearly using a Latin style.

A. Administrative

1. **Title:** *Proposal to encode 9 Latin letters for North American indigenous orthographies*
2. Requester's name: *...???..., Karl Pentzlin, ...???...*
3. Requester type (Member body/Liaison/Individual contribution): *TBD*
4. Submission date: *TBD*
5. Requester's reference (if applicable): *TBD*
6. Choose one of the following:
   This is a complete proposal: *Yes*
   (or) More information will be provided later: *No*

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters): *No*
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *Yes*
      Name of the existing block: *Latin Extended-D*
2. Number of characters in proposal: *10*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary   *X*   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
4. Is a repertoire including character names provided? *Yes*
      a. If YES, are the names in accordance with the "character naming guidelines"
         in Annex L of P&P document? *Yes*
      b. Are the character shapes attached in a legible form suitable for review? *Yes*
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for
   publishing the standard? *TBD*
   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools
   used:
6. References:
      a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? *Yes*
      b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
      of proposed characters attached? *Yes*
7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input,
   presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? *No*

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

---

[1] Form number: N3152-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? — *No*
    If YES explain

2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? — *TBD*
        If YES, with whom?
        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? — *TBD*
    Reference:

4. The context of use for the proposed characters (type of use; common or rare) — *common*
    Reference:

5. Are the proposed characters in current use by the user community? — *Yes*
        If YES, where? Reference: *TBD*

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? — *Yes*
            If YES, is a rationale provided? — *Yes*
                If YES, reference: *Contemporary use*

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? — *Yes*

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? — *No*
            If YES, is a rationale for its inclusion provided?
                If YES, reference:

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? — *No*
            If YES, is a rationale for its inclusion provided?
                If YES, reference:

10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? — *No*
            If YES, is a rationale for its inclusion provided?
                If YES, reference:

11. Does the proposal include use of combining characters and/or use of composite sequences? — *No*
        If YES, is a rationale for such use provided?
            If YES, reference:
        Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
            If YES, reference:

12. Does the proposal contain characters with any special properties such as control function or similar semantics? — *No*
            If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility character(s)? — *No*
        If YES, is the equivalent corresponding unified ideographic character(s) identified?
            If YES, reference: