L2/20-251

Unicode request for modifier Latin capital letters

Kirk Miller, kirkmiller@gmail.com

Craig Cornelius, ccornelius@google.com 2020 September 25

This is a request for modifier capital letters, two needed for Chatino orthography (Mexico) and the third for phonological use.

Of the basic Latin alphabet, Unicode does not provide modifier support for capital C F Q S X Y Z.

The Chatino community of Oaxaca (ca. 18,000 speakers as of 2000) has decided on an orthography that uses modifier capital ᴬ to ᴸ as tone letters. (See Figure 1.) Of these, ⟨ᶜ, ᶠ⟩ are missing.

Modifier capital letters are also used in more generic tone transcription, such as ⟨ᴴV ᴹV ᴸV ᴿV ᶠV⟩ for HIGH, MID, LOW, RISING and FALLING tonemes on a vowel syllable; FALLING is not supported by Unicode. (See Figure 9.)

Capital letters are commonly used as para-IPA wild cards for natural classes of sounds (e.g. C for 'consonant', N 'nasal', P 'plosive', F 'fricative', etc.), and are combined the same ways IPA letters are, including as modifiers. For example, the set of prenasalized consonants is {ᴺC}, a consonant with fricated release [Cᶠ], etc. Perhaps the most common use not supported by Unicode is ⟨Nᶜ, ᶜN⟩ for a generic post-stopped or pre-stopped nasal, or the set of such sounds.

Similarly with superscript letters for weak or incomplete articulation. The *Linguistic Atlas Project* for example specifies that ⟨ᶜC Cᶜ⟩ is the transcription for consonant sequences where one segment is weakly articulated. (See Figure 6.)

We also request ⟨ᵠ⟩ for use in Japanese phonology, where it is semantically distinct from ⟨Q⟩.

Modifier letters, capital

ᶜ U+A7F2 MODIFIER LETTER CAPITAL C. Figures 2–8, 11–13.

ᶠ U+A7F3 MODIFIER LETTER CAPITAL F. Figures 7–14.

ᵠ U+A7F4 MODIFIER LETTER CAPITAL Q. Figure 15.

Chart

As ⟨ᶜ, ᶠ⟩ are part of a community orthography, the BMP is requested for maximum support.

	…0	…1	…2	…3	…4	…5	…6	…7	…8	…9	…A	…B	…C	…D	…E	…F
Latin Extended-D																
U+A7Fx			ᶜ	ᶠ	ᵠ											

Properties

```
A7F2;MODIFIER LETTER CAPITAL C;Lm;0;L;<super> 0043;;;;N;;;;;
A7F3;MODIFIER LETTER CAPITAL F;Lm;0;L;<super> 0046;;;;N;;;;;
A7F4;MODIFIER LETTER CAPITAL Q;Lm;0;L;<super> 0051;;;;N;;;;;
```

References

Emiliana Cruz (2014) BSC mapa Quiahije. paisajechatino.wixsite.com/chatinolandscape/paisajes and Archive of the Indigenous Languages of Latin America, Chatino Landscape Collection, ailla.utexas.org/islandora/object/ailla:252384.

Emiliana Cruz, María Elena Méndez Cortés & Claudia García Baltazar (forthcoming 2020). Chatino prologue to the Spanish translation of Hernández Castillo, Hutchings & Noble (eds) *Transcontinental Dialogues*.

Isaura de los Santos. *Ykoa^{A} Mango^{HA} Panix* (*The Mango Tree*). StoryWeaver Community.

Osamu Fujimura & JC Williams (1999) "Syllable concatenators in Japanese, Spanish, and English", in Fujimura et al. (eds.) *Proceedings of LP '98: Item Order in Language and Speech*. 4th Linguistics and Phonetics Conference, Columbus, Ohio.

Nina Grønnum (2005, 2013) *Fonetik og Fonologi: Almen og dansk*. Akademisk forlag, Copenhagen.

H. A. G. Houghton (2016) *The Latin New Testament: A Guide to Its Early History, Texts, and Manuscripts*. Oxford University Press.

Patricia Keating, Daniel Wymark & Ryan Sharif (2019) "Proposal for superscript diacritics for prenasalization, preglottalization, and preaspiration," *JIPA*.

Renzo Miotti (2015) "Fonetica e fonologia," in Heinemann & Melchior (eds.) *Manuale di linguistica friulana*.

MUFI: Medieval Unicode Font Initiative (2015), 4th edition.

Lee Pederson (1986) *Linguistic Atlas of the Gulf States*, vol. 7. University of Georgia Press.

Martha Ratliff (2015) "Word-initial prenasalization in Southeast Asia," in Enfield & Comrie (eds.) *Languages of Mainland Southeast Asia*.

Anastasia Riehl & Abigail Cohn (2011) "Partially Nasal Segments," *The Blackwell Companion to Phonology*, vol. 1.

Tatsuo Takoo (2002) *A Catalogue and Index of the Shelley Manuscripts in the Bodleian Library and a General Index to the Facsimile Edition of the Bodleian Shelley Manuscripts*. Psychology Press.

Figures

Figure 1. Prof. Emiliana Cruz, a native speaker of Highlands Chatino, at a language workshop in San Juan Quiahije, Oaxaca. Cruz creates learning materials for local schools and libraries. The community orthography is seen here, though not the unsupported tone letters. (Source: Google Earth, accessed 2020 August 24.)

Modifier letter capital C (ᶜ)

Initial prenasalized stops are presented in grammars and language sketches variously as combinations of a syllabic nasal and a consonant (N.C), as consonant clusters (NC), as complex unitary stop phonemes (ᴺC), or as post-stopped nasals (Nᶜ). These descriptions vary with respect to transcription convention,

Figure 2. Ratliff (2015: 39). A semantic distinction between ⟨NC⟩, ⟨ᴺC⟩ and ⟨Nᶜ⟩ as wildcards.

Le occlu-costrittive non-sonore possono esser interessate dagli stessi fenomeni descritti per le occlusive (non-sonore), nelle medesime condizioni (parziale laringalizzazione della vocale accentata precedente, [ˈV²C], [VᶜˈCV]; cf. § 2.2).

Figure 3. Miotti (2015: 382)

Throughout the chapter, we have discussed the two types of partially nasal segments attested phonologically – prenasalized segments and prestopped nasals – in parallel, despite much greater documentation of ^{N}Cs than ^{C}Ns cross-linguistically. While we have acknowledged the paucity of information on the ^{C}N cases, we have not addressed the source of this imbalance. This imbalance raises several questions. First,

Figure 4. Riehl & Cohn (2011: 572). ⟨^{C}N⟩ for {prestopped nasal}.

prenasalized consonants. Ratliff (2015) describes the various historical developments of original prenasalized consonants in the languages of Mainland South East Asia, noting that across languages these consonants are now variably N̥C, NC, ^{N}C, N^{C}, or plain voiced C. Overall, then,

Figure 5. Keating, Wymark & Sharif (ms p. 8)

10. C^{C} indicating two consonants (or a vowel and a consonant, V^{C}) with the second member faintly articulated
11. ^{C}C indicating two consonants (or a vowel and a consonant, ^{V}C) with the first member faintly articulated

Figure 6. Pederson (1986: 29). Definition of ⟨^{C}⟩.

Figure 7. Cruz (2014). Place names in the Chatino municipality of San Juan Quiahije. One of the manually superscripted ^{F}'s has lost its formatting.

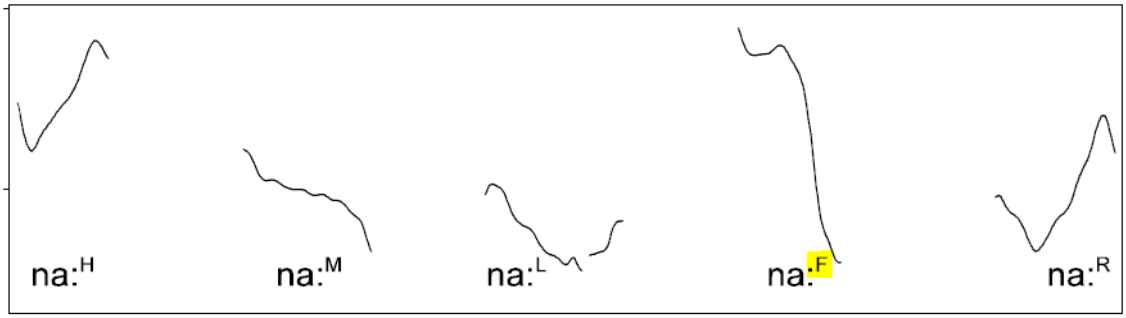
Cha?^{F} nga^{J} tyin^{H} ?in^{J} ktyi^{C} re^{C1}

Emiliana Cruz, María Elena Méndez Cortés, Claudia García Baltazar

Lo^{A} ktyi^{C} no^{E} nde^{C} na^{K} **cha?^{F} ndywen?^{G} ky?an^{J} s?en^{K}: ne^{J} ren?^{I} jnya^{F} ?o^{E} ne^{J} s?wa^{I} ren?^{A} ?o^{E} kchin^{A} jnya^{E} ?in^{J} Canada, Shya?^{A} ?o^{E} Australia^{2}** nten^{B} ?an^{E} no^{A} nya^{B} ?in^{E} ran^{F} nt?u^{B} ren?^{E} shka^{I} nya^{J} ?ne^{B} shla^{K} no^{A} ne^{J} sh?an^{H} ?in^{J} nten^{B}.^{3} Na^{F} no^{J} wa^{C} ngya^{G} na^{E} ren?^{J} ?o^{E} ?ne^{G} sh?an^{E} ren?^{J} ?in^{J} shla^{K} ?ne^{K} sh?an^{H} nten^{B} kchin^{A} ?in^{A} ne?^{A} pi^{H} no^{A} ykwi?^{A} ?in^{A} ran^{F} jlo^{J} (no^{A} lo^{I} la^{J} nga^{J} ne?^{A} pi^{H} nts?wi^{A} shka^{I} ?a^{A} ty?a^{A} t?o^{E}, ?o^{E} ne?^{A} pi^{H} no^{A} ndwi^{E} t?wa^{A} nt?an^{H} ?na^{G} ?o^{E}) shno^{H} ska^{J} na^{F} ?o^{E} cha?^{F} tkwi^{E} s?en^{A} no^{A} ti^{C} ndiya^{J} ni^{I} ?yan^{J} an^{I} ndiya^{A} jya?^{F} no^{A} yno^{E} na^{F} no^{A} ngwa^{C} ti^{C} s?ne^{E}, t?a^{J} nya^{E} ngwa^{C} na^{F} shi^{E} ?ne^{G} shla^{K}

Figure 8. Cruz et al. (forthcoming). The superscript digits are footnotes.

Modifier letter capital F (ᶠ)



Figur 9.9 De fem toner i thai

9.5.3 Notation
Der er flere forskellige måder at notere toner på i litteraturen. En ser således ud: ᴴpa – høj, ᴹpa – midt, ᴸpa – lav, ᶠpa – faldende, ᴿpa – stigende (‘*r*ising’). Meget almindeligt

Figure 9. Grønnum (2005: 201, 202). Modifier capitals as generic tone letters.

Figure 10. Website of the Chatino landscape (place name) project, paisajechatino.wixsite.com/chatinolandscape.

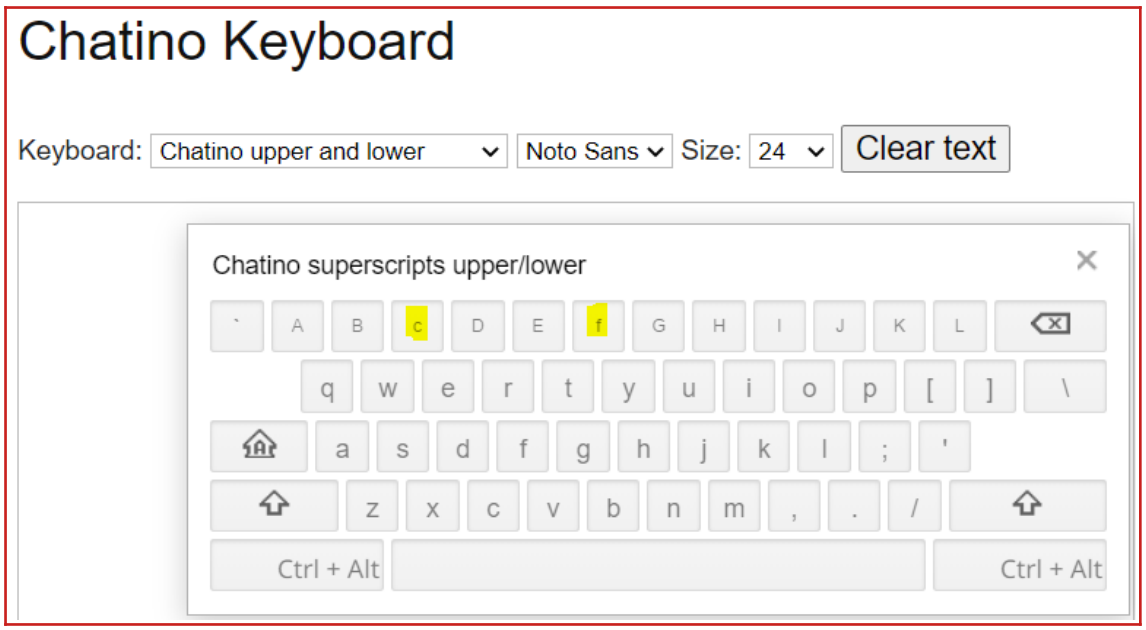


Figure 11. One of us (Cornelius) has created a Chatino keyboard, but resorts to modifier lower-case *c* and *f* for tone letters due to lack of Unicode support of the modifier capitals. languagetools-153419.appspot.com/omq/.

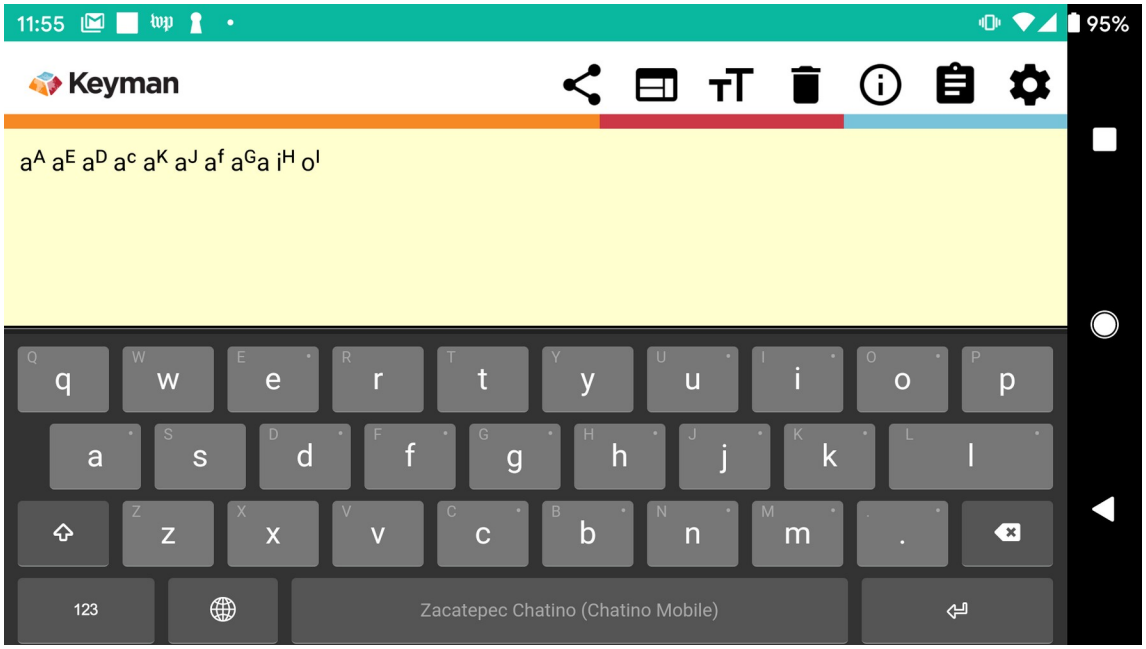


Figure 12. Sample of tone orthography generated by KeyMan Chatino input for Android, which is installed on phones in San Juan Quiahije. Long-press triggers the tone letters. Tones ^{C} and ^{F} are hacked with lower case.

Ndyqan^{A} stenq^{C} ndyqan^{A} na^{H} qnya^{H} jqen^{C}
tqwoa^{A} tyku^{E} tiyu^{H}.
ndyqan^{A} jyqan^{H} ndyqan^{A} na^{H} qnya^{H} jqen^{C}te^{A}
tqwoa^{A} neq^{C} lyqo^{F} qin^{E} qni^{A}. qa^{A} tkwen^{I} renq^{A}
nyka^{A} qin^{E} renq^{A} tyjoa^{C} renq^{A} qnya^{H}.

Figure 13. De los Santos (no date, p. 5). A primer for Chatino.

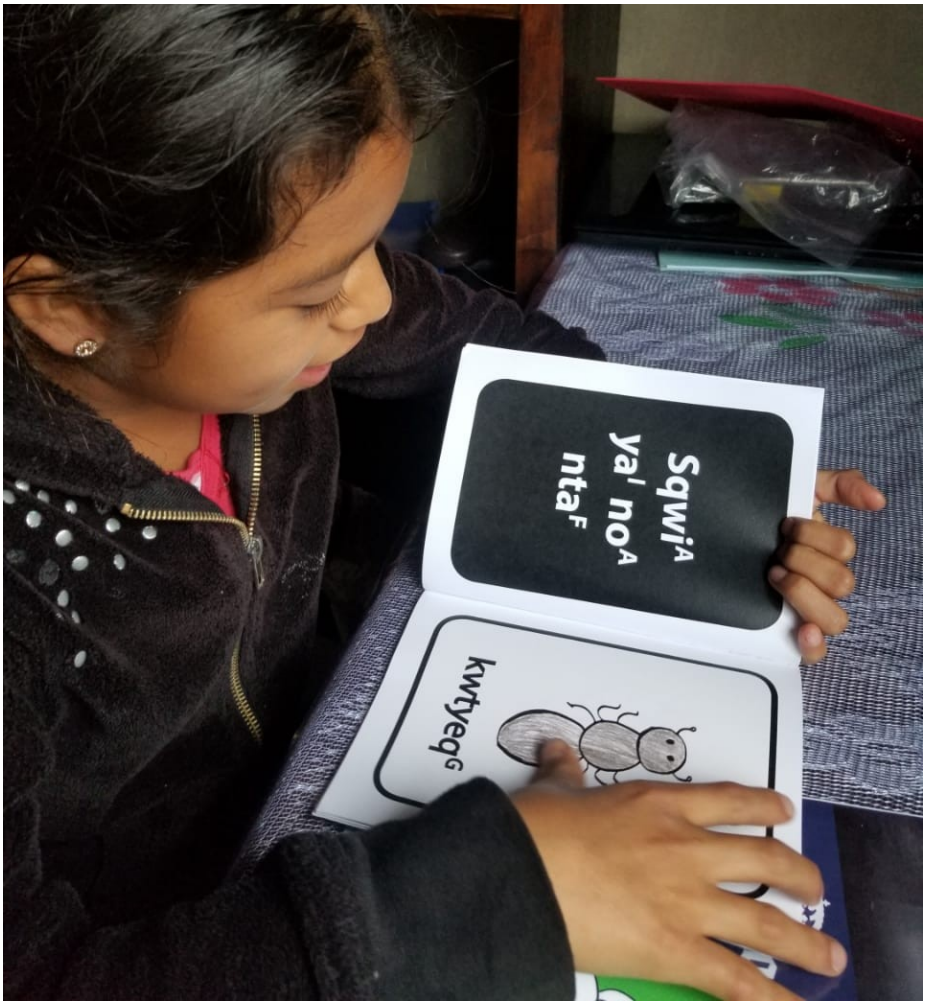
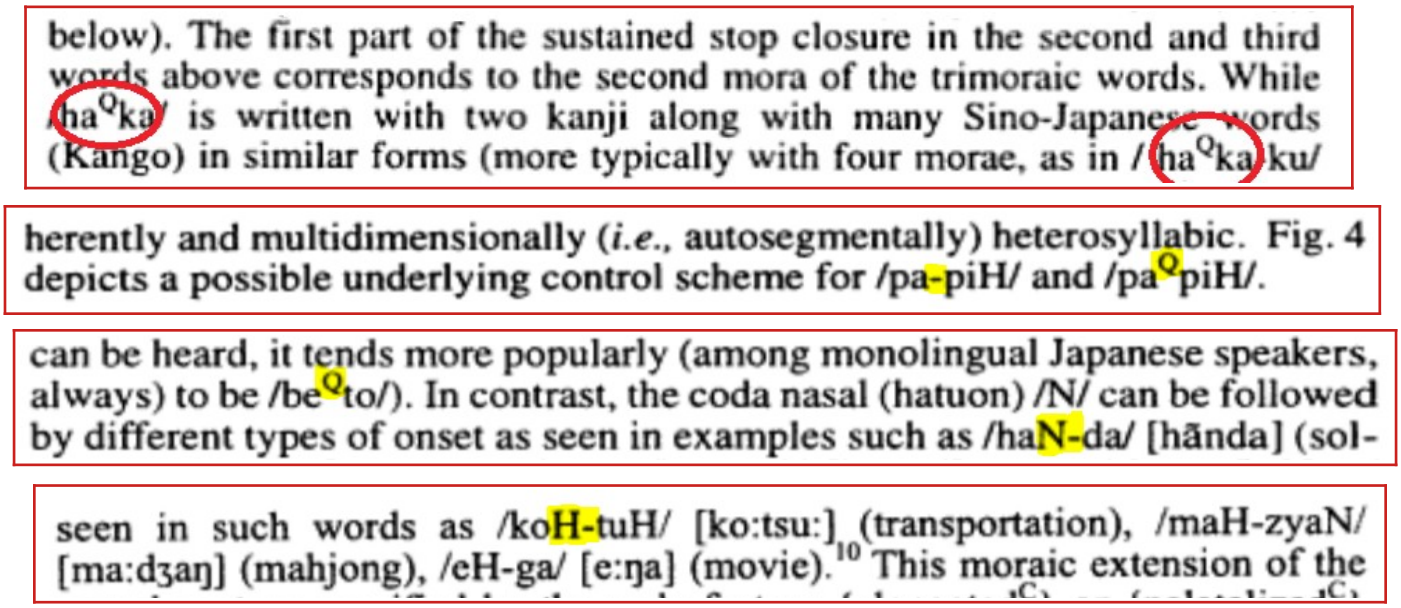


Figure 14. A child reading a Chatino primer with the unsupported tone letter ⟨F⟩. ichan.ciesas.edu.mx/lectura-en-chatino-en-tiempos-de-covid-19/.

Modifier letter capital Q (ᵠ)

Used for *sokuon* (phonemic gemination) in Japanese. Distinct from baseline Q.



below). The first part of the sustained stop closure in the second and third words above corresponds to the second mora of the trimoraic words. While /haᵠka/ is written with two kanji along with many Sino-Japanese words (Kango) in similar forms (more typically with four morae, as in /haᵠkaku/

herently and multidimensionally (*i.e.*, autosegmentally) heterosyllabic. Fig. 4 depicts a possible underlying control scheme for /pa-piH/ and /paᵠpiH/.

can be heard, it tends more popularly (among monolingual Japanese speakers, always) to be /beᵠto/). In contrast, the coda nasal (hatuon) /N/ can be followed by different types of onset as seen in examples such as /haN-da/ [hāⁿda] (sol-

seen in such words as /koH-tuH/ [ko:tsu:] (transportation), /maH-zyaN/ [ma:dʒaŋ] (mahjong), /eH-ga/ [e:ŋa] (movie).[10] This moraic extension of the

Figure 15. Fujimura & Williams (1999: 473, 481, 474, 475) specify (p. 474) that at syllable boundaries they "use superscript Q for sokuon in place of the regular syllable boundary hyphen" (which they use when there is no sokuon, as in /haᵠka-ku/ circled in red, or /pa-piH/ vs /paᵠpiH/ in yellow). This use of superscript ⟨ᵠ⟩ for sokuon as a syllable concatenator contrasts with traditional baseline ⟨Q⟩ for sokuon as a segment, and contrasts here with baseline ⟨N⟩ and ⟨H⟩ for moraic nasal and vowel length as segments.

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
See also http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** *Modifier Latin capital letters*
2. Requester's name: *Kirk Miller, Craig Cornelius*
3. Requester type (Member body/Liaison/Individual contribution): *individual*
4. Submission date: *2020 September 25*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *yes*
   (or) More information will be provided later:

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters):
      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: *3*
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. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? *Kirk Miller*
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): *SIL (Gentium release)*
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)? *yes*

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 Unicode Character Database (http://www.unicode.org/reports/tr44/) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

1 Form number: N4502-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, 2009-11, 2011-03, 2012-01)

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.)? *yes*
   If YES, with whom? *Emiliana Cruz (Prof. Anthropology, CIESAS-CDMX), Chatino Landscape Project*
   If YES, available relevant documents: *[see illustrations]*
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? *yes*
   Reference: *Ethnologue reports on Highland Chatino (ISO codes ctp, cly, cya)*
4. The context of use for the proposed characters (type of use; common or rare) *phonetic, orthographic*
   Reference:
5. Are the proposed characters in current use by the user community? *yes*
   If YES, where? Reference: *Oaxaca, Mexico. See illustrations.*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? *no*
   If YES, is a rationale provided?
   If YES, reference:
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? *at least [C], [F]*
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, or could be confused with, 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 characters? *no*
    If YES, are the equivalent corresponding unified ideographic characters identified?
    If YES, reference: