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, \(\langle c, f \rangle\) are missing.

Modifier capital letters are also used in more generic tone transcription, such as \(\langle \text{HIGH}\text{ MID}\text{ LOW}\text{ RISING}\text{ FALLING} \rangle\) 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 \(\{\text{N}C\}\), a consonant with fricated release \(\text{C}^{\text{F}}\), etc. Perhaps the most common use not supported by Unicode is \(\langle \text{N}^{\text{C}}, \text{C}^{\text{N}} \rangle\) 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 \(\langle C^{\text{C}} \rangle\) is the transcription for consonant sequences where one segment is weakly articulated. (See Figure 6.)

We also request \(\langle q \rangle\) for use in Japanese phonology, where it is semantically distinct from \(\langle Q \rangle\).

**Modifier letters, capital**

- \(\text{C}\) \(\text{U+A7F2 MODIFIER LETTER CAPITAL C. Figures 2–8, 11–13.}\)
- \(\text{F}\) \(\text{U+A7F3 MODIFIER LETTER CAPITAL F. Figures 7–14.}\)
- \(\text{Q}\) \(\text{U+A7F4 MODIFIER LETTER CAPITAL Q. Figure 15.}\)

**Chart**

As \(\langle c, f \rangle\) 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  | Q  |    |    |    |    |    |    |    |    |    |    |    |    |
Properties

References


Isaura de los Santos. Ykoaᴬ Mangoᴴᴬ Panix (The Mango Tree). StoryWeaver Community.


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 occluso-costitutive non-sonore possono esser interessate dagli stessi fenomeni descritti per le occlusive (non-sonore), nelle medesime condizioni (parziale laringalizzazione della vocale accentata precedente, [V:\C], [V:\C\V]; 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 NCs than NCs cross-linguistically. While we have acknowledged the paucity of information on the NC cases, we have not addressed the source of this imbalance. This imbalance raises several questions. First,

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 NC, NC, ÑC, Ñ̃C, or plain voiced C. Overall, then,

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

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

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

Figure 8. Cruz et al. (forthcoming). The superscript digits are footnotes.
Modifier letter capital F (F)

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.
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 ⟨↑⟩.  
Modifier letter capital Q (♩)

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

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 /ha♩ka-ku/ circled in red, or /pa♩piH/ vs /pa♩piH/ in yellow). This use of superscript ♩ 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.
A. Administrative

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

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: |
| 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) |
| C-Major extinct x D-Attested 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. |

## C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?
   - **No**

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**
   - **Emiliana Cruz (Prof. Anthropology, CIESAS-CDMX), Chatino Landscape Project**

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?
   - **Yes**
   - **[see illustrations]**

4. The context of use for the proposed characters (type of use; common or rare)
   - **Phonetic, orthographic**

5. Are the proposed characters in current use by the user community?
   - **Yes**
   - **Reference:**

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?
   - **No**

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

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?
   - **No**

9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?
   - **No**

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**

11. Does the proposal include use of combining characters and/or use of composite sequences?
    - **No**

12. Does the proposal contain characters with any special properties such as control function or similar semantics?
    - **No**

13. Does the proposal contain any Ideographic compatibility characters?
    - **No**

---

**References:**
- Ethnologue reports on Highland Chatino (ISO codes ctp, cly, cya)
- Oaxaca, Mexico. See illustrations.