

Unicode request for modifier capital S

Kirk Miller, kirkmiller, gmail

2024 January 30

This proposal is for a modifier capital ‘S’, <^s>, which is used both as a phonetic symbol and as an orthographic letter. We request that it be added to the BMP.

<^s> is used as a phonetic and phonemic wildcard. When capital <S> is the wildcard for ‘fricative’ or ‘sibilant’, modifier <^s> may indicate a fricative or sibilant release, or may be combined with the wildcard for ‘plosive,’ together denoting an affricate, for example <T^s> in Figure 4. <^s> is also attested as a spirantization phoneme, as in Figure 3, similar to other modifier capital letters for phonemes without a specific phonetic realization, such as <^q> for gemination in the Japanese tradition that was adopted with L2/20-251.

<^s> is also used as a tone-sandhi letter in the orthography of Western Highlands Chatino of Oaxaca (Oto-Manguan, ISO [CTP]). The modern Chatino orthography was designed for the San Juan Quiahije dialect, and there is increasing adoption in communities that speak the neighboring Yaitepec and Panixtlahuaca dialects (Hilaria Cruz, p.c. 2022). These three dialects, which together constitute Western Highlands Chatino, had traditionally used Spanish orthography. Chatino orthography uses a number of modifier capital letters for lexical tone and tone sandhi, with <^s> the single gap in Unicode support. Unicode 14 added the unsupported lexical-tone letters <^{c f}> to the Latin Extended-D block, following request [L2/20-251](#). Of the letters used for tone sandhi, only <^s> is unsupported. <^s> is requested by Chatino-speaker Hilaria Cruz of the University of Louisville and her colleagues Francis Tyers of Indiana University and Emiliana Cruz of CIESAS-CDMX.

Character

^s A7F1 MODIFIER LETTER CAPITAL S. Figures 4–5.

Properties

A7F1;MODIFIER LETTER CAPITAL S;Lm;0;L;<sup> 0053;;;N;;;;

Chart

We request <^s> be placed in the Latin Extended-D block. During the SAH discussion of L2/20-251, Michael Everson suggested that all modifier capital letters of the basic Latin alphabet should be placed in the BMP, and that suggestion was followed with <^c>, <^f> and <^q>, even though <^q> was not needed for orthography.

| | A720 | | | Latin Extended-D block | | | | | | A7FF | | | | |
|---|------|-----|-----|------------------------|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|
| | A72 | A73 | A74 | A75 | A76 | A77 | A78 | A79 | A7A | A7B | A7C | A7D | A7E | A7F |
| 0 | Ƨ | F | K | P | W | 9 | Ƨ | N | € | X | Ø | □ | | |
| 1 | Ǝ | s | k | p | w | ɖ | l | n | g | L | ø | □ | | s |
| 2 | 3 | AA | K | P | z | ɭ | n | € | K | J | W | B | | C |
| 3 | 3 | aa | k | p | z | mɭ | n | € | k | X | w | ß | | F |
| 4 | ‘ | AO | K | Ɔ | Ɔ | nɭ | Ƨ | ç | N | B | Ç | ß | | Q |
| 5 | ‘ | æ | k | Ɔ | Ɔ | rɀ | Ƨ | hɁ | n | ß | ç | ß | | I |
| 6 | H | AJ | L | Q | P | Rɀ | Ƨ | B | R | Ω | ZɁ | □ | | ı |
| 7 | h | ai | l | q | p | t | Ƨ | b | r | ω | Ð | □ | | ı |
| 8 | Ƨ | A | L | Q | Ƨ | ɀ | ^ | f | s | U | d | δ | | н |
| 9 | Ƨ | a | l | q | Ƨ | Ɔ | : | f | s | u | S | δ | | œ |
| A | ε | Æ | Θ | Ƨ | 3 | ɖ | = | Ƨ | H | ʼA | s | λ | | III |
| B | ε | æ | θ | Ƨ | 3 | Ƨ | ! | a | 3 | á | Ƨ | λ | | Ƨ |
| C | 4 | A | Ɔ | Ƨ | Ƨ | Ƨ | ' | Ƨ | g | ʼI | Ƨ | λ | | Ƨ |
| D | 4 | a | Ɔ | Ƨ | Ƨ | Ƨ | Ƨ | Ƨ | Ƨ | ı | Ƨ | | | W |
| E | 4 | Ƨ | ∞ | Ƨ | 9 | Ƨ | Ƨ | Ƨ | Ƨ | ʼU | Ƨ | | | I |
| F | 4 | Ƨ | ∞ | Ƨ | 9 | Ƨ | · | Ƨ | Q | ı | Ƨ | | | W |

References

- Hilaria Cruz (2017). Orthographic manuscript to 'Prayers for the Community: Parallelism and Performance in San Juan Quiahije Eastern Chatino.' *Oral Tradition* 31/2: 509–534.
- Ricardo “Mende” Cruz Cruz (2004). Speech addressed to City Helpers in San Juan Quiahije on December 31, 2004. Recorded by Hilaria Cruz.
- Stacey Inez Oberly (2008) *A Phonetic Analysis of Southern Ute with a Discussion of Southern Ute Language Policies and Revitalization*. University of Arizona.
- Vennemann (1985) The bifurcation theory of the Germanic and German consonant shifts. In Fisiak (ed.) *Papers from the 6th International Conference on Historical Linguistics*. John Benjamins, Amsterdam. 527–547.

Figures

Chatino

⟨^s⟩ indicates an extra-high floating tone in Western Highlands Chatino. This tone /^s/, which is higher in pitch than the lexical high tone /^k/, is a product of tone sandhi and is very common in verbal inflections. For example, when the noun phrase /yka^A tyi^F nta^K/ is combined with the verb phrase /ndywiq^A renq^J qo^E janq^G/, the high tone of nta^K changes the tone of the verb ndywiq^A ‘3sg said’ from low /^A/ to extra-high /^s/: “Yka^A tyi^F nta^K ndywiq^S renq^J qo^E janq^G” ‘They call it a ykатыinta tree’ (Hilaria Cruz, pc. 2022).

| | | | |
|---|------------------------------------------------------------------------|--------------------------------------------------------|--------------------------|
| 1 | <i>ndya^I-ra^Ksqi^S</i> | <i>qa^J ykwenq^Jen^I</i> | sometimes |
| 4 | | | wrongly |
| | | | we spoke |
| 1 | <i>Kwiq^J Ndyo^B-si^S</i> | <i>janq^G</i> | Just God was |
| 6 | | | |
| 1 | <i>kanq^G no^A nke^G qna^G</i> | | the one who |
| 7 | | | was |
| | | | patient with us |
| 1 | <i>kanq^G no^A nda^F</i> | <i>jwe^A-sa^S</i> | the one who gave: |
| 8 | <i>qna^G</i> | | strength to us |

Figure 1. Cruz (2004, lines 14, 16–18). Tone letter ⟨^s⟩ in the transcription of a public speech. Line 14 shows a contrast between the high lexical tone /^k/ and the extra-high floating tone /^s/. The colon in line 18 indicates prosodic lengthening.

| Text 2 Prayer for the Community performed by Wenceslao Cortés. | | | | |
|----------------------------------------------------------------|----|------------------------------------------|-------------------------------------|---------------------------------------------|
| Part I | | | | |
| 1 | I | Ka ^J tyi ^H | ton ^K ne ^S , | Now they are about to rise up, |
| | | <small>JUST POT:AUX.TO STAND NOW</small> | | |
| 2 | I | Ka ^J tyi ^H | tqen ^J ne ^C . | Now they are about to establish themselves. |
| | | <small>JUST POT:AUX.TO EXIST NOW</small> | | |
| 3 | II | Ka ^J kqu ^S | ni ^S , | Now they are about to grow, |
| | | <small>JUST POT:GROW NOW</small> | | |
| 4 | II | Ka ^B klu ^S | ni ^S , | Now they are about to thrive. |
| | | <small>JUST POT:THRIVE NOW</small> | | |
| 5 | II | Ka ^J tykwa ^S | ni ^S , | Now they are about to sit up, |
| | | <small>JUST POT:SIT.ELEVATED NOW</small> | | |
| 6 | II | Ka ^J tyqin ^I | ni ^C . | Now they are about to establish themselves. |
| | | <small>JUST POT:EXIST NOW</small> | | |

Figure 2. Cruz (2017, appendix 2). An unusually long sequence containing tone /^s/. /ni^C/ ‘now’ becomes /ni^S/ after tone /^k/ or another tone /^s/. (⟨Ka^B⟩ lost its HTML formatting in line 4.)

Phonological symbol

FIGURE (6) Timbisha Final Feature (Miller et. al. 2005: 415)

| <i>/-paʔa/, 'on'</i> | | | | | |
|----------------------|-------------------------------|-------------|----------------|----------------|-----------------------|
| a. Gemination | <i>/tua^G-/</i> | 'son' | + | <i>/-paʔa/</i> | → <i>[tuappaʔa]</i> |
| b. Nasalization | <i>/pijin^N-/</i> | 'duck' | + | <i>/-paʔa/</i> | → <i>[pijimbaʔa]</i> |
| c. Spirantization | <i>/naka^S-/</i> | 'big horn'+ | <i>/-paʔa/</i> | → | <i>[naʔaβaʔa]</i> |
| d. Aspiration | <i>/haintfi^H-/</i> | 'friend' | + | <i>/-paʔa/</i> | → <i>[haintʃɸaʔa]</i> |

Figure 3. Oberly (2008: 38). Phonemic modifier letters: /^G/ gemination, /^N/ nasalization, /^S/ spirantization and /^H/ aspiration. Here /^Sp/ is realized as [β]. (Baseline capitals /G N S H/ would indicate phonemic segments in their own right rather than modification of an adjacent segment.)

The essential parts of the developments, those involving the assumed PIE plosives, are summarized in (1), where capital letters are used for entire series of plosives, fricatives, and affricates, e.g. T for all voiceless fortis plosives, þ for all voiceless fortis fricatives (except *s*), Ð for all voiceless lenis fricatives (except *z*), **T^s** for all affricates. As Grimm put

(8) The High Gmc. frication

T^s → +TS → +SS (→ +S)

Figure 4. Vennemann (1985: 527, 535). The <^s> is identified as a capital in the text; baseline capital <S> is the wildcard for fricatives (see p. 541), alongside <T> for voiceless/fortis plosives. The historical evolution of the High Germanic frication summarized here is ancestral T^s [affricate] → TS [plosive-fricative sequence] → SS [geminate fricative] → S [simple fricative].

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹.

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from 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 std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html.
See also std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html for latest Roadmaps.

A. Administrative

| | |
|------------------------------------------------------------------|---------------------------|
| 1. Title: | <i>Modifier capital S</i> |
| 2. Requester's name: | <i>Kirk Miller</i> |
| 3. Requester type (Member body/Liaison/Individual contribution): | <i>individual</i> |
| 4. Submission date: | <i>2024 January 30</i> |
| 5. Requester's reference (if applicable): | |
| 6. Choose one of the following: | |
| This is a complete proposal: | <i>yes</i> |
| (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: | <i>yes</i> |
| Name of the existing block: | <i>Latin Extended-D</i> |
| 2. Number of characters in proposal: | <i>1</i> |
| 3. Proposed category (select one from below - see section 2.2 of P&P document): | |
| A-Contemporary <input checked="" type="checkbox"/> | B.1-Specialized (small collection) <input type="checkbox"/> |
| C-Major extinct <input type="checkbox"/> | D-Attested extinct <input type="checkbox"/> |
| F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/> | E-Minor extinct <input type="checkbox"/> |
| | G-Obscure or questionable usage symbols <input type="checkbox"/> |
| 4. Is a repertoire including character names provided? | <i>yes</i> |
| a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document? | <i>yes</i> |
| b. Are the character shapes attached in a legible form suitable for review? | <i>yes</i> |
| 5. Fonts related: | |
| a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? | <i>Kirk Miller</i> |
| b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): | <i>SIL (Gentium release)</i> |
| 6. References: | |
| a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? | <i>yes</i> |
| b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? | <i>yes</i> |
| 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)? | <i>no</i> |

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? | Hilaria Cruz (University of Louisville), Francis Tyers (Indiana University) |
| If YES, available relevant documents: | p.c. emails available |
| 3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? | ca. 26,000 |
| Reference: | 16,000 Western Highland Chatino: Instituto Nacional de Lenguas Indígenas (2000) |
| 4. The context of use for the proposed characters (type of use; common or rare) | literacy material |
| Reference: | |
| 5. Are the proposed characters in current use by the user community? | yes |
| If YES, where? Reference: | See references |
| 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: | use for basic literacy material |
| 7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? | (with Chatino) |
| 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: | |