L2/21-041

Unicode request for additional para-IPA letters

Kirk Miller, kirkmiller@gmail.com 2021 January 11

This is a supplement to L2/20-125R ‘Unicode request for expected IPA retroflex letters and similar letters with hooks’, which was accepted by the UTC in 2020 July.

These three characters derive from the common convention of ⟨c⟩ and ⟨č⟩ as affricates [ts] and [tš]. This convention has been extended by re-purposing IPA ⟨ɕ⟩ as an alveolo-palatal *affricate* and creating a new letter ⟨ȼ⟩ for the retroflex [tʂ]. Because ⟨ɕ⟩ as an affricate conflicts with its IPA definition as a fricative, a new letter ⟨ꞩ⟩ was created for the fricative, along with modifier ⟨ˢ⟩. Thus, in this system, ⟨c č ȼ ɕ⟩ correspond to [ts tš tʂ tɕ].

A capital ⟨Ȼ⟩ is expected as an IPA wildcard for {retroflex consonant}, but it is not yet attested and is only noted here for reference.

Characters

- ˢ 107BA MODIFIER LETTER SMALL S WITH CURL. Figures 9–10.
- ȼ 1DF1D LATIN SMALL LETTER C WITH RETROFLEX HOOK. Figures 1–5, 8.
- ꞩ 1DF1E LATIN SMALL LETTER S WITH CURL. Figures 1–2, 6–8.

Properties

```
107BA;MODIFIER LETTER SMALL S WITH CURL;Lm;0;L;<super> 1DF1E;
    ;;;N;;;;;
1DF1D;LATIN SMALL LETTER C WITH RETROFLEX HOOK;Ll;0;L;;;;;N;;;;;
1DF1E;LATIN SMALL LETTER S WITH CURL;Ll;0;L;;;;;N;;;;;
```

Chart

Characters on a grey background have been accepted by the UTC.

	...0	...1	...2	...3	...4	...5	...6	...7	...8	...9	...A	...B	...C	...D	...E	...F
Latin Extended-F																
U+107Bx	ˠ	ˣ	ˤ	ˀ	ˁ	˙	˛	˚	˜	˝	ˢ					
Latin Extended-G																
U+1DF1x	ᴤ	ᴦ	ʤ	ʧ	ɳ	ɻ	ʂ	ʧ	ʒ	ʤ	ᶖ	ᶗ	ᶯ	ȼ	ꞩ	

References

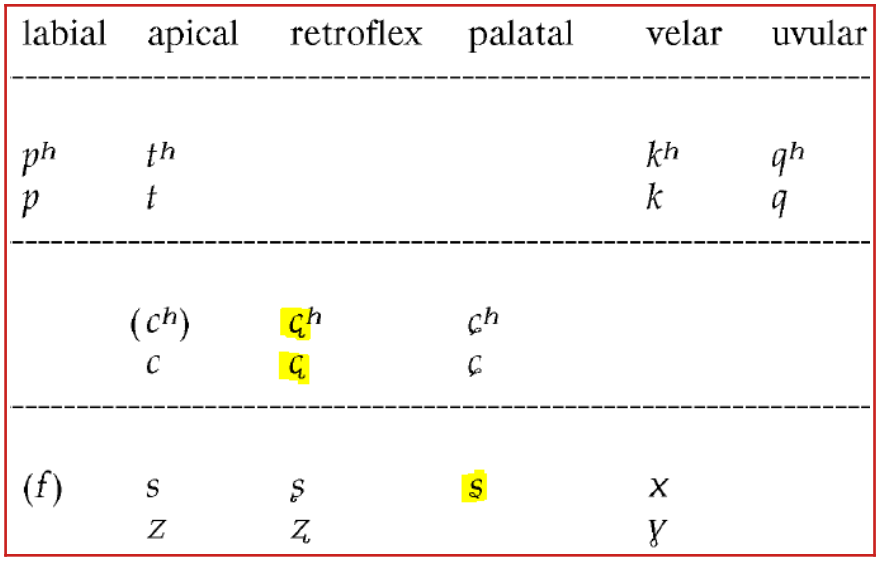
John Kelly & John Local (1989) *Doing Phonology: Observing, Recording, Interpreting.* Manchester University Press.

Martina Roos (2000) *The Western Yugur (Yellow Yugur) Language*. PhD thesis, Leiden University.

Hans Nugteren & Marti Roos (1998) 'Common vocabulary of the Eastern and Western Yugur languages: the Tibetan loanwords', *Studia Etymologica Cracoviensia* 3: 45–92s.

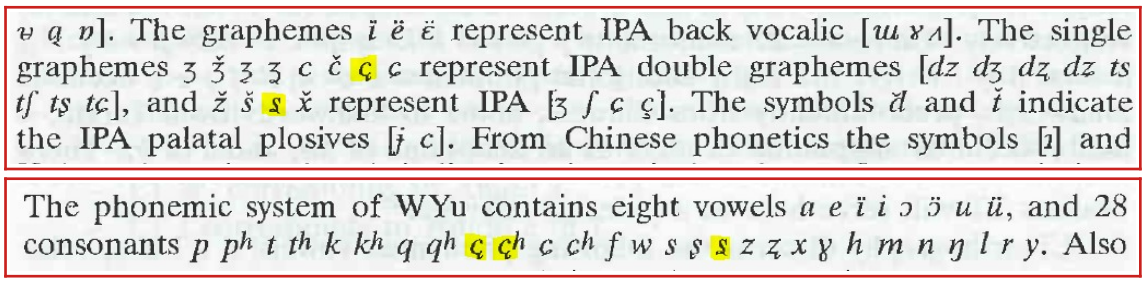
Figures

Preview



labial	apical	retroflex	palatal	velar	uvular
p^h	t^h			k^h	q^h
p	t			k	q
	(c^h)	$ʈ^h$	$ɕ^h$		
	c	$ʈ$	$ɕ$		
(f)	s	$ʂ$	$ʃ$	x	
	z	$ʐ$		$ɣ$	

Figure 1. Roos (2000: 18). Identification of ⟨ʈ⟩. and ⟨ʂ⟩.



ʉ ɑ ɒ]. The graphemes ï ë ë represent IPA back vocalic [ɯ ɤ ʌ]. The single graphemes ʒ ǰ ʒ ʒ c č ʈ ɕ represent IPA double graphemes [dz dʒ dʐ dz ts tʃ tʂ tɕ], and ž š ʃ x̌ represent IPA [ʒ ʃ ɕ ʂ]. The symbols đ and ť indicate the IPA palatal plosives [ɟ c]. From Chinese phonetics the symbols [ɻ] and

The phonemic system of WYu contains eight vowels *a e ï i ɔ ɔ̈ u ü*, and 28 consonants *p pʰ t tʰ k kʰ q qʰ ʈ ʈʰ c cʰ f w s ʂ ʃ z ʐ x ɣ h m n ŋ l r y*. Also

Figure 2. Nugteren & Roos (1998: 49).

Small c with retroflex hook (ɕ)

2.4.10. The phoneme /ɕ/ is realized as a voiceless retroflex affricate [ɕ].

The phoneme /ɕ/ occurs in the following positions in the word:
initial: ***ɕal=*** 'to pray', ***ɕowat=*** 'to thank'.
intervocalic: ***aɕï=*** 'to be worried', ***quɕaqta=*** 'to hug', ***yaŋqïɕa*** 'alone', ***qaʰɕan*** 'when'.
cluster (ɕ second member): ***ɕaʰpɕïɣïr*** 'hoe', ***aʰtɕï*** 'horseherd', ***ɕʰïmïkɕï*** 'thief', ***poʰqɕa*** 'bag', ***kusɕï*** 'cowherd', ***ɕʰamɕa*** 'shirt', ***törtïnɕï*** 'fourth', ***qïlɕïqʰ-qalɕïqʰ*** 'crooked and bent', ***eʰlɕï*** 'shaman', ***temïrɕï*** 'smith', ***qoyɕï*** 'shepherd'.

2.4.11. The phoneme /ɕʰ/ is realized as a voiceless aspirated retroflex affricate [ɕʰ], and occurs in initial position only.

Examples of minimal pairs /ɕ/ - /ɕʰ/:

ɕotta= 'to scold'	***ɕʰotta=*** 'to libate'
ɕïl 'sheep droppings'	***ɕʰïla=*** 'to be angry'
ɕalïm 'hoar frost'	***ɕʰamɕa*** 'shirt'
ɕoqï= 'to pinch'	***ɕʰoqe=*** 'to sit'

Figure 3. Roos (2000: 22). Explanation of ⟨ɕ⟩.

[7] WYu *ɕʰïlwï* C82:70b, *ɕïlwu* L226b, *ɕïlwï* L285b, *ɕïɬwɔ* T177a 'bell'; EYu *čėlßė* B144 'id'. From LT dril-bu Das655a 'id', Arïk *ɕėr-wė* TB666, Zêkog *pɕė-wė* Qú91:167, Xiàhé *ɕėl-wė* HL283b, ET *dril-wu* Gō1177 'id'. Cf Mgr *ʂilu*: S/M88, *ʂirḅu* S/M92, *ɕiliu* Khas221 (cf RT136), BaoÑ *ɕlwė* CN209 'id'. The Arïk and Zêkog forms are peculiar.

Figure 4. Nugteren & Roos (1998: 62).

105. *antaqanta-ya pu aɕa-sï pïr qara-ɣanta pu yiɣït sueyʐan kʰïr-tï tanʂï taʰqï-ya ɕïta-ttï*[6].
106. *'men qïztar-nï saɣa per-in' ti-ɣintï.*
107. *antaqanta-ya pu amsa kuan-nïŋ qïz-ï-tï.*
108. *'am saɣa per-si sen teylï ɕɛnyü* [..] *ɕʰoqe-ɣïmistï, un-ïk ki-ʂtï.'*

Figure 5. Roos (2000: 188). Examples of ⟨ɕ⟩ in textual transcription.

Small s with curl (ʂ)

2.4.16. The phoneme /ʂ/ is realized as a voiceless front-prepalatal fricative [ʂ].

The phoneme /ʂ/ does not occur frequently: apart from Chinese loanwords, there are only six instances of initial /ʂ/: ***ʂiki*** 'slender', ***ʂor=*** 'to sweep', ***ʂoz*** 'word', ***ʂozïr=*** 'to drag', ***ʂoɣïl=*** 'to bake', ***ʂünan*** 'Sùnán', and two instances in which /ʂ/ occurs as the second member of a cluster: ***kanʂü*** 'Gānsù', ***kunsin*** ~ ***kunʂin*** 'every day'.

Example of a minimal pair /s/ - /ʂ/:
sor= 'to suck' ***ʂor=*** 'to sweep'

Example of a (near) minimal pair /s/ - /ʂ/ - /ʂ/:
siki 'goat' ***ʂiki*** 'slender' ***ʂïkï*** 'two'

Figure 6. Roos (2000: 23). Explanation of ⟨ʂ⟩.

ʂüɛʂo (~ ***ʂüɛ***) 'school', cf L245b, C82:70b, *ʂöʂo* L245a, *ʂïuse* T211a, *ʂüe* L245a, *ʂö* L245a.
¶ ←Ch xuéxiào, xué.
° Also loanword in EYu *ʂüeʂiao* B112.
ʂünan 'Sùnán (county)', cf *sunan* CL95.
¶ ←Ch sùnán; for the inexplicable front-prepalatal ʂ, see also ***kanʂü*** 'Gānzhōu'.

Figure 7. Roos (2000: 366). Examples of ⟨ʂ⟩.

Typical of Arïk, Zêkog, Huari, and Prževal'skij, is the retention of a labial element before the newly developed affricates in first syllables, as in Arïk *wʂæ* 'bird', *wʂok* 'direction', *pçaŋ* 'chest' from LT bya, phyog, braṅ[35].

Figure 8. Nugteren & Roos (1998: 53).

Modifier s with curl (ˢ)

The phonemic system of WYu contains eight vowels *a e ï i ɔ ö u ü*, and 28 consonants *p pʰ t tʰ k kʰ q qʰ ç çʰ c cʰ f w s ʂ ʃ z ʐ x ɣ h m n ŋ l r y*. Also operating within the system as a distinguishing feature is the phenomenon of preaspiration, realized as an aspiration element *ʰ*, or allophonically as a fricative *χ*, *φ*, or *ˢ*; preaspiration and subsequent *l* is realized as *ɬ*.

Figure 9. Nugteren & Roos (1998: 49). Explanation of [ˢC] as an allophone of /ʰC/.

When following a high front vowel, preaspiration may be realized as a voiceless front-prepalatal fricative [ˢʂ], e.g. ***iʰt*** 'meat' [i̥ˢʂt̚], ***yiʰp*** 'rope' [ʐiˢʂp̚], ***iʰçikï*** 'intestines' [i̥ˢʂciɡ̊ï], ***piʰkʊ*** 'very' [pi̥ˢʂko], ***iʰçinti*** 'in it' [i̥ˢʂcin̥di], ***iʰçʊtï*** 's/he drinks' [i̥ˢʂco̥d̥ï], ***yiʰtï*** 'I ate' [ʐi̥ˢʂtï], ***üʰçüz*** 'three hundred' [ÿˢʂüz]. In ***töʰrçin*** 'village; chief', preaspiration is realized as [törˢʂin].

Figure 10. Roos (2000: 33). Examples of [ˢC].

Expected character

Capital C with retroflex hook (C̢)

⟨C̢⟩ is the expected wildcard letter for the set of {retroflex consonants}, analogous to ⟨Cʼ⟩ for {ejective consonant(s)} and ⟨Ƈ⟩ for {implosive consonant(s)}. It is not yet attested.

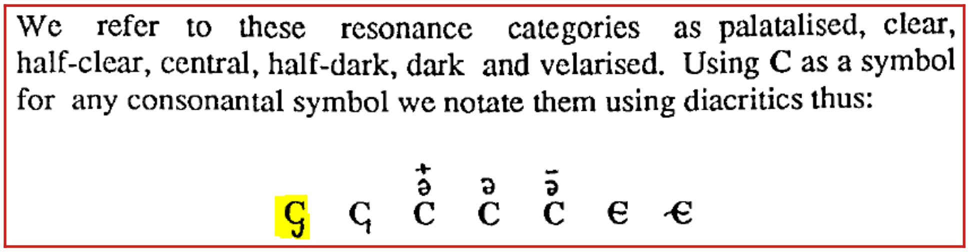

We refer to these resonance categories as palatalised, clear, half-clear, central, half-dark, dark and velarised. Using C as a symbol for any consonantal symbol we notate them using diacritics thus:

Figure 11. Kelly & Local (1989: 73), illustrating the common convention of modifying capital ⟨C⟩ with diacritics for various sets of consonants. Highlighted U+A7C4 ⟨Ꞔ⟩ at bottom left, with the old IPA convention of a left hook for {palatalized consonant}, is analogous to expected ⟨C̢⟩ for {retroflex consonant}.

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:** *Additional phonetic click letters*
2. Requester's name: *Kirk Miller*
3. Requester type (Member body/Liaison/Individual contribution): *individual*
4. Submission date: *2021 January 11*
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: *x*
   (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: *x*
      Name of the existing block: *Latin Extended-F, G*
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? *The author is a members of the user community.*
   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?
   Reference:
4. The context of use for the proposed characters (type of use; common or rare) *phonetic*
   Reference:
5. Are the proposed characters in current use by the user community? *yes*
   If YES, where? Reference: *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)? *no*
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? *yes*
   If YES, is a rationale for its inclusion provided?
   If YES, reference: *(Unicode disprefers use of combining retroflex and palatal hooks)*
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? *no*
    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: