L2/21-043

Unicode request for subscript modifier-letter support

Kirk Miller, kirkmiller@gmail.com 2021 January 11

Of lower case subscript basic Latin letters, Unicode currently lacks *b c d f g q w y z*. Of subscript Greek, Unicode supports ᵦ ᵧ ᵨ ᵩ ᵪ. The only additional subscript of an IPA letter is schwa ₔ.

In IPA transcription, some authors restrict superscript modifiers to sequential sounds (onsets, on/off-glides, releases and weakly or incompletely articulated sounds), transcribing simultaneous articulation instead with subscripts. This partially follows UPA conventions. An example is Yuen Ren Chao (1968) *Language and Symbolic Systems*, p. 32 ff, which uses subscript *n r j*. Penhallurick (1991) uses subscript ᵻ ᶎ ᵣ ᵦ for r-coloring of a preceding vowel, as opposed to a superscript, which indicates a weakened rhotic consonant. In some Americanist traditions as well, superscript and subscript are distinguished, but so far I've only found subscript *a e i o u*, which are already supported by Unicode.

In diplomatic (as opposed to normalized) transliteration of abugidas that have conjuncts (i.e., the placement of one letter under another to indicate the lack of an intervening vowel), a Latin subscript may be used for the second element of the conjunct. Cambodian, for example, would have subscript *b c d f ñ ŋ q w y*, though I have yet to find Cambodian manuscripts in diplomatic transliteration, and thus cannot attest to subscript *ŋ* or *q*. Subscript *ñ* (seen in the examples) and *ç* should presumably be handled as *n* and *c* plus a diacritic.

For Americanist notation, the American Anthropological Association (*Phonetic Transcription Of Indian Languages*, 1916, p. 9) says, "If it is desired to distinguish between vocalic timbres and weakly articulated voiceless vowels on the one hand and vocalic glides and weakly articulated voiced vowels on the other hand, superior vowels (ᵃ) may be used for the former, inferior vowels (ₐ) for the latter." The letters that are not supported in subscript are open ɔ and Greek *α ε ι υ ω* (ₔ ᵅ ᵋ ᵢ ᵤ ᵚ). (So far, only ⟨ₑ⟩ has been attested.) In SIL fonts, most Americanist combining diacritics appear to work properly on subscripts, or at least on some of them: dieresis ⟨ä ₐ̈⟩, single dot ⟨ȧ ₐ̇⟩ and vertical stroke ⟨ḁ ₐ̍⟩, though the ogonek needs work: ⟨ą ₐ̨⟩.

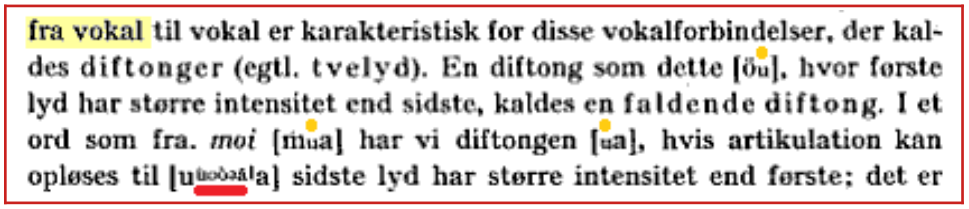

fra vokal til vokal er karakteristisk for disse vokalforbindelser, der kaldes diftonger (egtl. tvelyd). En diftong som dette [öu], hvor første lyd har større intensitet end sidste, kaldes en faldende diftong. I et ord som fra. *moi* [mua] har vi diftongen [ua], hvis artikulation kan opløses til [u̯o̯ə̯a̯la] sidste lyd har større intensitet end første; det er

Figure 1. Hammerich (1934: 59). This snippet view is not particularly clear, but shows weak elements of diphthongs printed in small script (yellow), as was common at the time. The Spanish RFE phonetic alphabet uses this convention, though only ⟨đ⟩ is attested. These small letters are simply typographic variants of subscripts and should be encoded as such.

Characters

Latin and IPA

ᵨ 1DF20 LATIN SUBSCRIPT SMALL LETTER TURNED ALPHA. Figure 24.

ᵦ 1DF21 LATIN SUBSCRIPT SMALL LETTER B. Figures 2–3.

ᶜ 1DF22 LATIN SUBSCRIPT SMALL LETTER C. Figures 2–3, 5–6.

ᵈ 1DF23 LATIN SUBSCRIPT SMALL LETTER D. Figures 2–4.

ᶞ 1DF24 LATIN SUBSCRIPT SMALL LETTER D WITH STROKE. Figure 7.

ᶿ 1DF25 LATIN SUBSCRIPT SMALL LETTER ETH. Figure 27.

ᵋ 1DF26 LATIN SUBSCRIPT SMALL LETTER OPEN E. Figure 27.

ᶠ 1DF27 LATIN SUBSCRIPT SMALL LETTER F. Figure 27.

ᵧ 1DF29 LATIN SUBSCRIPT SMALL LETTER GAMMA. Figures 8–10.

ᵩ 1DF2A LATIN SUBSCRIPT SMALL LETTER PHI. Figure 28.

ᵣ 1DF2B LATIN SUBSCRIPT SMALL CAPITAL R. Figures 11, 14.

ᵨ 1DF2C LATIN SUBSCRIPT SMALL LETTER TURNED R. Figures 11, 14.

ᵲ 1DF2D LATIN SUBSCRIPT SMALL CAPITAL INVERTED R. Figures 11–12.

ᵳ 1DF2E LATIN SUBSCRIPT SMALL LETTER R WITH TAIL. Figures 11–12.

ᶳ 1DF2F LATIN SUBSCRIPT SMALL LETTER ESH. Figure 25.

ᵤ 1DF30 LATIN SUBSCRIPT SMALL CAPITAL U. Figure 27.

ᵩ 1DF31 LATIN SUBSCRIPT SMALL LETTER UPSILON. Figure 24.

ᵥ 1DF32 LATIN SUBSCRIPT SMALL LETTER TURNED V. Figures 26–27.

ʷ 1DF33 LATIN SUBSCRIPT SMALL LETTER W. Figures 15–16, 18.

ʸ 1DF34 LATIN SUBSCRIPT SMALL LETTER Y. Figures 17, 19.

ᶻ 1DF35 LATIN SUBSCRIPT SMALL LETTER Z. Figures 20–23.

Greek

ᵤ 0379 GREEK SUBSCRIPT SMALL LETTER MU. Figure 28.

Chart

	...0	...1	...2	...3	...4	...5	...6	...7	...8	...9	...A	...B	...C	...D	...E	...F
Greek and Coptic																
U+037x									μ	μ						
Latin Extended-G																
U+1DF1x																
U+1DF2x	ɒ	b	c	d	đ	ð	ɛ	f	g	ɣ	ɸ	ʀ	ɹ	ʁ	ɽ	ʃ
U+1DF3x	ᴜ	ʊ	ʌ	w	y	z										

Properties

```
0379;GREEK SUBSCRIPT SMALL LETTER MU;Lm;0;L;<sub> 03BC;;;;N;;;;;
1DF20;LATIN SUBSCRIPT SMALL LETTER TURNED ALPHA;Lm;0;L;<sub> 0252;;;;N;;;;;
1DF21;LATIN SUBSCRIPT SMALL LETTER B;Lm;0;L;<sub> 0062;;;;N;;;;;
1DF22;LATIN SUBSCRIPT SMALL LETTER C;Lm;0;L;<sub> 0063;;;;N;;;;;
1DF23;LATIN SUBSCRIPT SMALL LETTER D;Lm;0;L;<sub> 0064;;;;N;;;;;
1DF24;LATIN SUBSCRIPT SMALL LETTER D WITH STROKE;Lm;0;L;<sub> 0111;;;;N;;;;;
1DF25;LATIN SUBSCRIPT SMALL LETTER ETH;Lm;0;L;<sub> 00F0;;;;N;;;;;
1DF26;LATIN SUBSCRIPT SMALL LETTER OPEN E;Lm;0;L;<sub> 025B;;;;N;;;;;
1DF27;LATIN SUBSCRIPT SMALL LETTER F;Lm;0;L;<sub> 0066;;;;N;;;;;
1DF28;LATIN SUBSCRIPT SMALL LETTER G;Lm;0;L;<sub> 0067;;;;N;;;;;
1DF29;LATIN SUBSCRIPT SMALL LETTER GAMMA;Lm;0;L;<sub> 0263;;;;N;;;;;
1DF2A;LATIN SUBSCRIPT SMALL LETTER PHI;Lm;0;L;<sub> 0278;;;;N;;;;;
1DF2B;LATIN SUBSCRIPT SMALL CAPITAL R;Lm;0;L;<sub> 0280;;;;N;;;;;
1DF2C;LATIN SUBSCRIPT SMALL LETTER TURNED R;Lm;0;L;<sub> 0279;;;;N;;;;;
1DF2D;LATIN SUBSCRIPT SMALL CAPITAL INVERTED R;Lm;0;L;<sub> 0281;;;;N;;;;;
1DF2E;LATIN SUBSCRIPT SMALL LETTER R WITH TAIL;Lm;0;L;<sub> 027D;;;;N;;;;;
1DF2F;LATIN SUBSCRIPT SMALL LETTER ESH;Lm;0;L;<sub> 0283;;;;N;;;;;
1DF30;LATIN SUBSCRIPT SMALL CAPITAL U;Lm;0;L;<sub> 1D1C;;;;N;;;;;
1DF31;LATIN SUBSCRIPT SMALL LETTER UPSILON;Lm;0;L;<sub> 028A;;;;N;;;;;
1DF32;LATIN SUBSCRIPT SMALL LETTER TURNED V;Lm;0;L;<sub> 028C;;;;N;;;;;
1DF33;LATIN SUBSCRIPT SMALL LETTER W;Lm;0;L;<sub> 0077;;;;N;;;;;
1DF34;LATIN SUBSCRIPT SMALL LETTER Y;Lm;0;L;<sub> 0079;;;;N;;;;;
1DF35;LATIN SUBSCRIPT SMALL LETTER Z;Lm;0;L;<sub> 007A;;;;N;;;;;
```

References

Andersen (1977) "Dansk Dialektologi i Tiåret 1965–74," *Danske Studieren*.

Pedro Martín Butragueño (2014) *Fonología variable del español de México*. Vol. I Procesos segmentales.

Juan Lope Blanch, ed. (1990) *Atlas lingüístico de México*. Vol. I, Fonética. Centro de Estudios Lingüísticos y Literarios, El Colegio de México, Fondo de Cultura Económica.

Bethin (1998) *Slavic Prosody*.

Bickford & Floyd (2006) *Articulatory Phonetics*.

Boas, Goddard, Sapir & Kroeber (1916) *Phonetic Transcription of Indian Languages: Report of Committee of American Anthropological Association* (Smithsonian Miscellaneous Collections 66.6)

Burquest (2009) *Análisis fonológico*.

Charbonneau & Jacques (1972) "[t_s] et [d_z] en français canadien," in Valdman (ed.) *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*.

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

Louis Leonor Hammerich (1934) *Tysk fonetik*

Hickey (2014) *The Sound Structure of Modern Irish*.

Patricia Keating (2018) 'Some challenges for the IPA from languages of China'. 13th Phonetic Conference of China, Jinan University, Guangzhou.

Kozok, Uli (2015) *A 14th Century Malay Code of Laws*. ISEAS, Singapore.

Lass, Norman (2012) *Contemporary Issues in Experimental Phonetics*. Elsevier.

Ling Feng (2007) 'The articulatory and acoustic study of fricative vowels in Suzhou Chinese'. *Proceedings of ICPhS XVI*, International Congress of Phonetic Sciences.

Pulleyblank (1962) "The Consonantal System of Old Chinese," *Asia Major* 9: 2.

RFE (1915) 'Alfabeto fonético de la revista de filología española', *Revista de filología española*, vol. II.

Skjærvø (2002) *Khotanese Manuscripts from Chinese Turkestan in the British Library*, part 2, vol. 5.

Tatham & Morton (2016) *Speech Production and Perception*.

Wickelgren, Wayne (1969) 'Context sensitive coding, associative memory, and serial order in (speech) behavior'. *Psychological Review*, 76 (1).

Figures

Subscript b c d g (b c d g)

Used for conjuncts in diplomatic transliteration of alphasyllabaries such as Javanese. A subscript thus indicates a different glyph or phonetic realization than the full letter, and must be retained for accurate transliteration. (A similar convention uses superscript vowels for *plene* writing of adjads.) Subscript *b c d w* are attested. Subscript *g ŋ* should occur, as they transliterate letters of the Javanese script that have conjunct (subscript) forms, but so far I have not found them in texts. Subscript *f q* occur in Javanese in foreign loans, and they also have conjunct forms. (*Q* is also a common convention in the area for glottal stop.) Subscript *ñ* occurs in Javanese, but can be handled by the font: ⟨$_{ñ}$⟩. See also illustration subscript *w* above; the *y z* found below are also expected as conjuncts in Javanese.

Subscripting in the diplomatic transliteration is not predictable from the normalized transliteration. For example, in Kozok (2015: P/5/6), normalized *geggah rabutti rampassi* is diplomatic ⟨gaggaḥ rabut$_t$irampas$_s$i⟩, with neither the second *g* in *gg* nor the *p* in *mp* subscript.

Subscript *b d g* are also used for unreleased stops in Chinese.

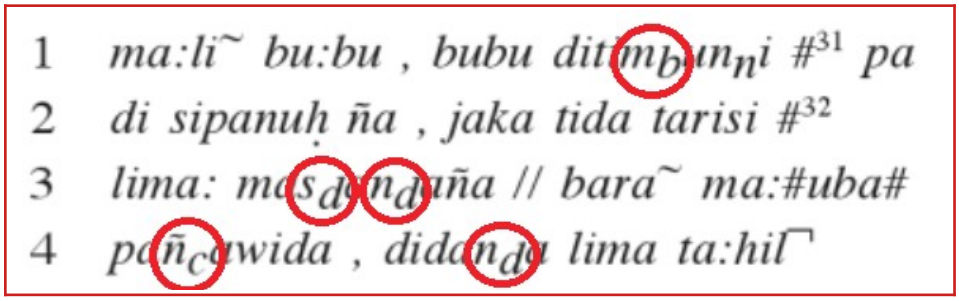


Figure 2. Kozok (2015: 69). The conjuncts commonly occur after nasals, but note ⟨s$_d$, p$_d$⟩ here and next, also ⟨k$_b$⟩ on p. 67.

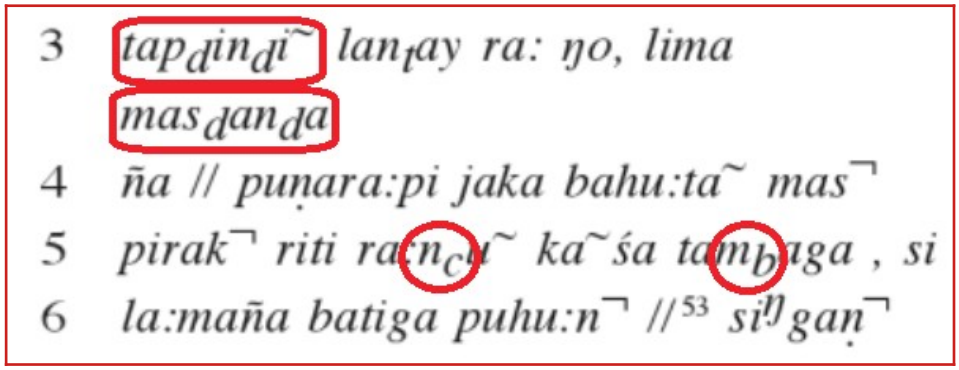


Figure 3. Kozok (2015: 71)

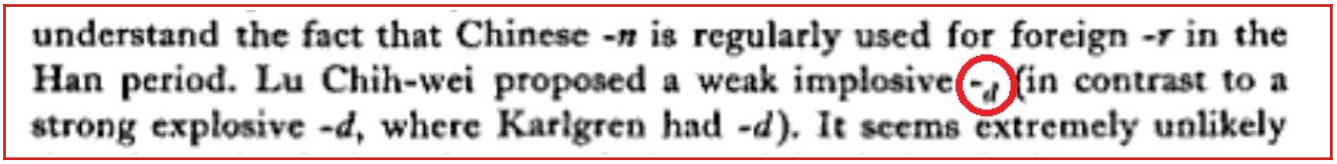
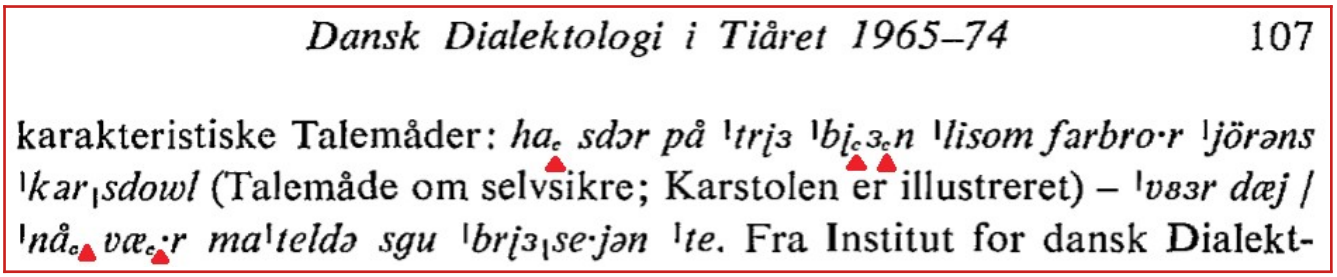
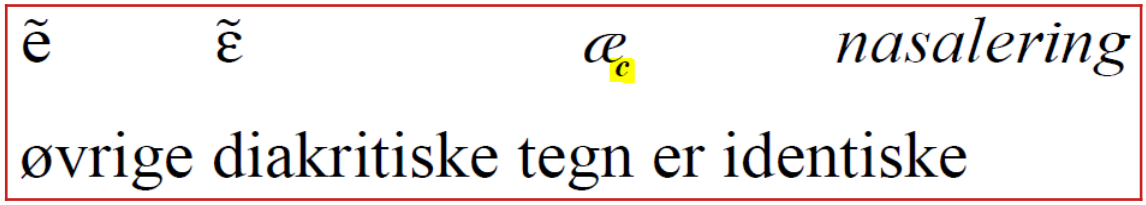

understand the fact that Chinese *-n* is regularly used for foreign *-r* in the Han period. Lu Chih-wei proposed a weak implosive *-d* (in contrast to a strong explosive *-d*, where Karlgren had *-d*). It seems extremely unlikely

Figure 4. Pulleyblank (1962: 215). Later on the page Pulleyblank says that all three final plosives appear to have been implosive. ("Implosive" here seems to mean unreleased.) In this notation the three would be $^{-}_{b}$, $^{-}_{d}$, $^{-}_{g}$.


Dansk Dialektologi i Tiåret 1965–74 107

karakteristiske Talemåder: *haˌ sdɔr på ˈtrɟɜ ˈbiˌɕɜˌn ˈlisom farbro·r ˈjörəns ˈkarˌsdowl* (Talemåde om selvsikre; Karstolen er illustreret) – *ˈvɛɜr dæj / ˈnåˌ væˌːr maˈteldə sgu ˈbrɟɜˌseˑjən ˈte*. Fra Institut for dansk Dialekt-

Figure 5. Andersen (1977:107). This 'nasal curl', semantically equivalent to an ogonek, is typeset in the Dania font that has been shared with me as a subscript 'c', and this published example appears to be a subscript 'c' as well.


ẽ ɛ̃ æ*c* *nasalering*

øvrige diakritiske tegn er identiske

Figure 6. Grønnum (2005: 419). In this publication as well, the 'nasal curl' is typeset as a subscript 'c'.

Subscript đ (đ)

The traditional Spanish phonetic alphabet (the RFE, or *Revista de filología española*, alphabet) uses reduced letters, equivalent to subscripts, for reduced sounds. ⟨đ⟩ in particular is provided for; I have not seen others. The RFE alphabet is used and taught today in Spanish-language universities, notably in Spain and Mexico. Apart from reduced sounds, it is covered by Unicode.

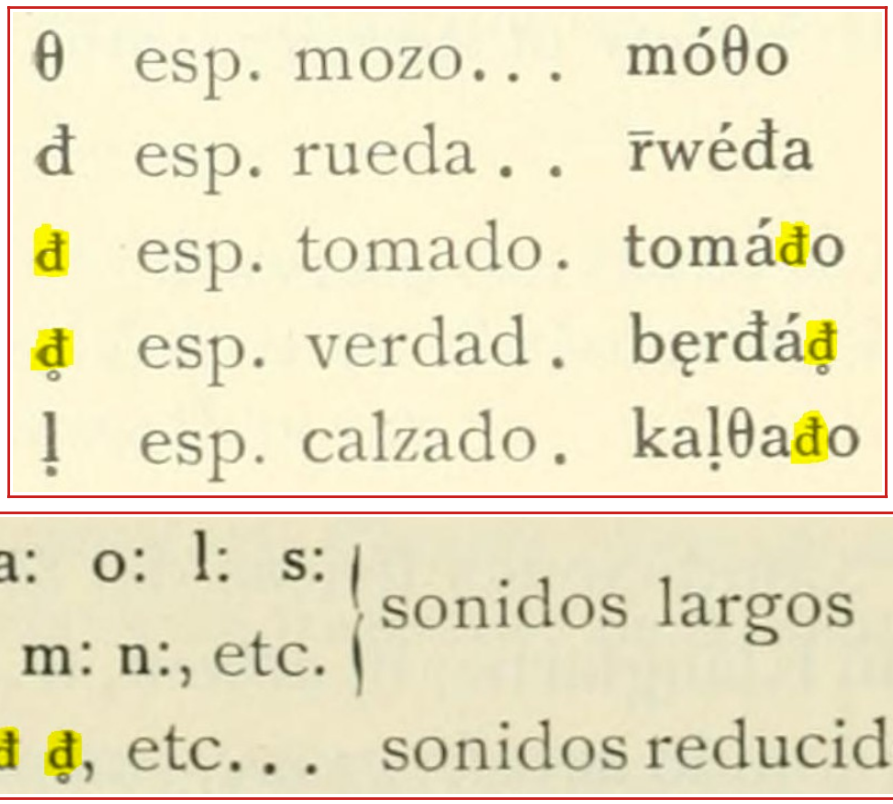


Figure 7. RFE (1915: 374–375). Contrast between normal ⟨đ⟩, small ⟨$_{đ}$⟩ and small voiceless ⟨$_{ḍ}$⟩, in the consonant table. The RFE alphabet chart has been reproduced many times over the past century, including in 2020 by Alexander Iribar of the phonetics laboratory at the University of Deusto in the Spanish Basque country. The second clip shows that reducing font size to indicate phonetic reduction is a productive convention, but I've only seen ⟨đ⟩ used in this way.

Subscript gamma (ᵧ)

Used for simultaneous velarization as opposed to a velar offglide (Bickford & Floyd) or relatively weak velarization (Hickey).

In the Irish of Roscommon/East Galway[25], Ring and Cape Clear voiced sibilants have been reported as the outcome of nasalising /s/, e.g. *i Sasana* [ɪ zasənˠə] 'in England'.

Figure 8. Hickey (2011) *The Dialects of Irish*, p. 31. Subscript notation is described next.

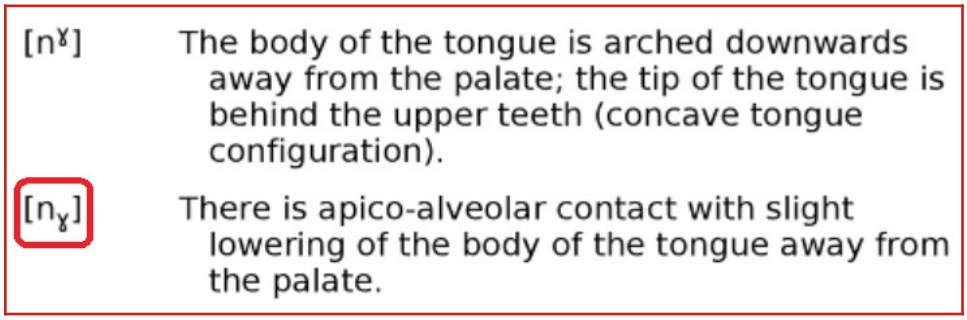


Figure 9. Hickey (2014: list just before §1.8.5; no page number in ebook). In the paragraph before §1.2 Hickey explains that a subscript gamma is used for southern dialects that have weak velarization. There is a similar distinction between [nʲ] (palatal) and [nⱼ] (palatalized alveolar).

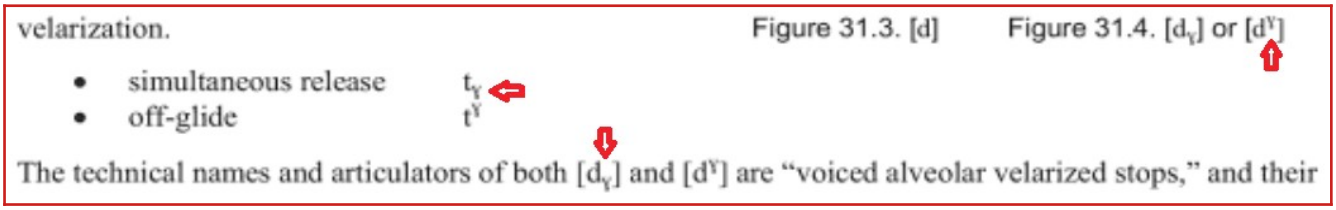


Figure 10. Bickford & Floyd (2006: 162)

Subscript ɹ ɽ ʀ ʁ (ᵢ ᵣ ʀ ʁ)

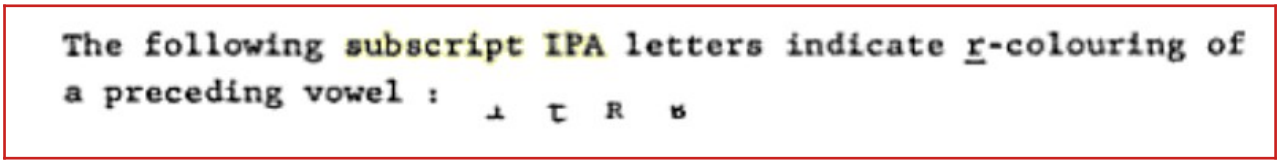
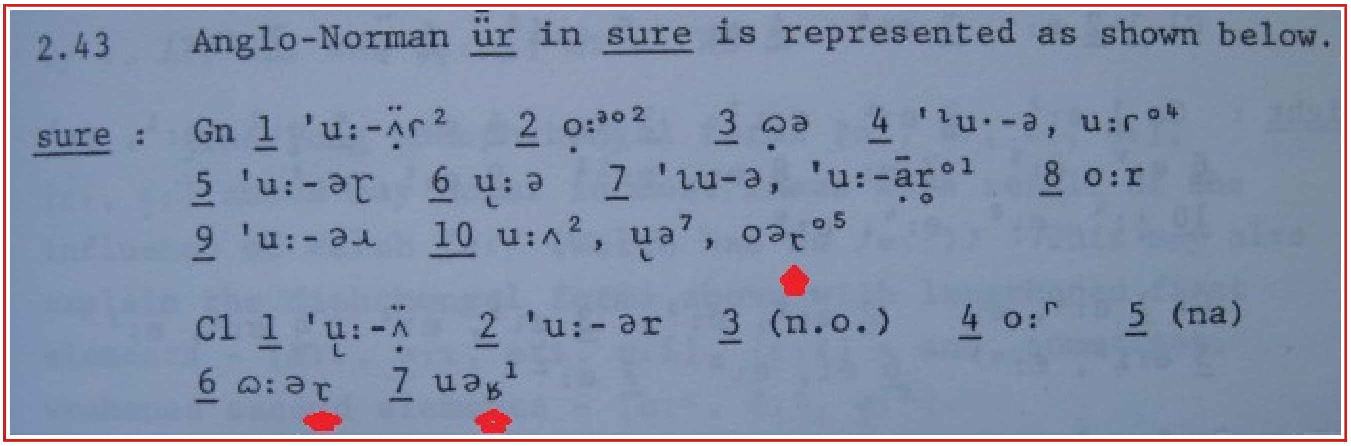


Figure 11. Penhallurick (1991) *The Anglo-Welsh Dialects of North Wales*, p. xviii. Subscript ɹ ɽ ʀ ʁ are used in the data.

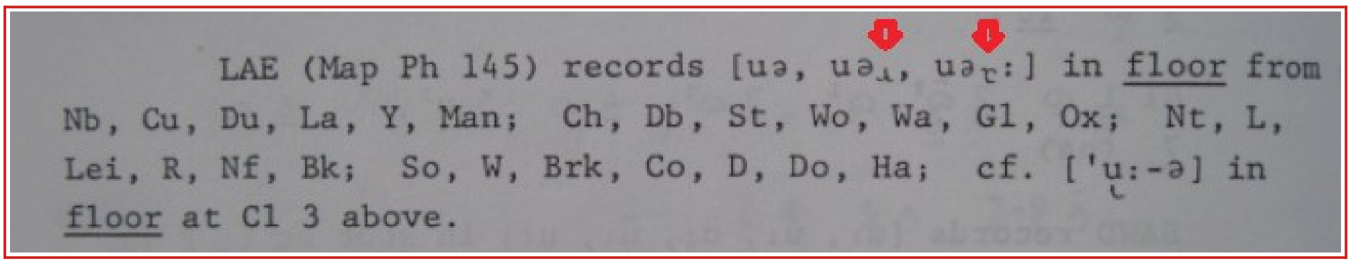


2.43 Anglo-Norman ür in sure is represented as shown below.

sure : Gn 1 'u:-ʌ̈ɾ[2] 2 o̤:ᵊᵒ[2] 3 ɷ̣ə 4 'ˡu·-ə, u:ɾ[o4]
5 'u:-əɽ 6 u̢:ə 7 'ɩu-ə, 'u:-ār̥[o1] 8 o:r
9 'u:-əɹ 10 u:ʌ[2], u̢ə[7], oəɽ[o5]

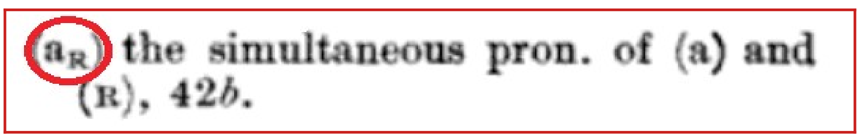
Cl 1 'u̢:-ʌ̈ 2 'u:-ər 3 (n.o.) 4 o:ʳ 5 (na)
6 ɷ:əɽ 7 uəʁ[1]

Figure 12. Penhallurick (1991: 85). Subscript *ɽ ʁ* along with superscript *ɾ.*



LAE (Map Ph 145) records [uə, uəɹ, uəɽ:] in floor from Nb, Cu, Du, La, Y, Man; Ch, Db, St, Wo, Wa, Gl, Ox; Nt, L, Lei, R, Nf, Bk; So, W, Brk, Co, D, Do, Ha; cf. ['u̢:-ə] in floor at Cl 3 above.

Figure 13. Penhallurick (1991: 85). Subscript *ɹ ɽ.* I have not found subscript *ʀ* in this volume, but it is used in the data it was drawn from (see first figure above). The data is published in volume 3 of D.R. Parry (director, U. of Swansea) & Penhallurick (ed.) *Survey of Anglo-Welsh Dialects*, which I don't have access to.



(aʀ) the simultaneous pron. of (a) and (ʀ), 42*b*.

Figure 14. Ellis (1889) *On Early English Pronunciation*, part V, p. 78.

Subscript w and y ($_{w\ y}$)

Americanist usage, occasional IPA usage per above, and used for the transcription of conjuncts of abugida scripts.

Como un segundo ejemplo de labialización obsérvense los datos del angas en (161). En estos datos, el símbolo [$_{w}$] pequeño debajo de ciertas consonantes representa el redondeamiento de los labios durante la articulación de la consonante sin una fuerte labialización en la distensión de ésta. La letra sobre escrita [^{w}] es una distensión labializada normal de la consonante:

(161)	/po/	[p$_{w}$o]	'boca'
	/bum/	[b$_{w}$um]	'gorro'
	/tu/	[t$_{w}$u]	'matar'
	/du/	[d$_{w}$u]	'oler'
	/ko/	[k^{w}o]	'o'

Figure 15. Burquest (2009: 130)

Labialized consonants, that is, consonants pronounced with simultaneous lip-rounding, are to be indicated by means of inferior *w* closely following the character. Thus, l_w indicates an *l* pronounced with markedly rounded lips; similarly, k_w indicates a *k* with simultaneous lip-rounding (not to be confused, of course, with *kw*).

Figure 16. Boas, Goddard, Sapir & Kroeber (1916: 15)

Palatalized consonants, that is, consonants modified by the simultaneous articulation of a large part of the surface of the tongue against the palate (in other words, by the tongue taking *y*-position), are to be indicated by closely following inferior *y*. Thus, n_y indicates a palatalized dental *n*. The ordinary so-called "palatal" *l* and *n* are probably best considered as palatalized dorsal *l* and *n* and should thus, strictly speaking, be indicated by λ_y (Italian *gl*) and ν_y (Italian *gn*); l_y and n_y would, however, be the normal methods of representing these consonants.

Figure 17. Ibid. (subsequent paragraph to previous figure)

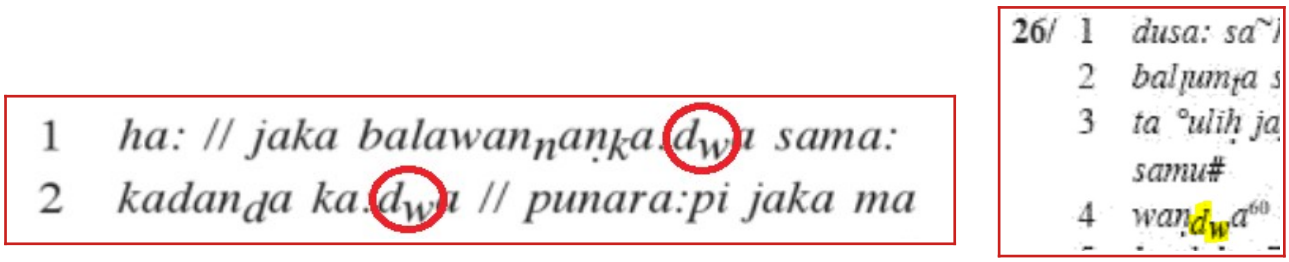


Figure 18. Kozok (2015: 71). A letter may be twice subscripted for a double conjunct (e.g. p. 72, line 26/4), but that is not semantically distinct from a simple conjunct and so perhaps does not require Unicode support.

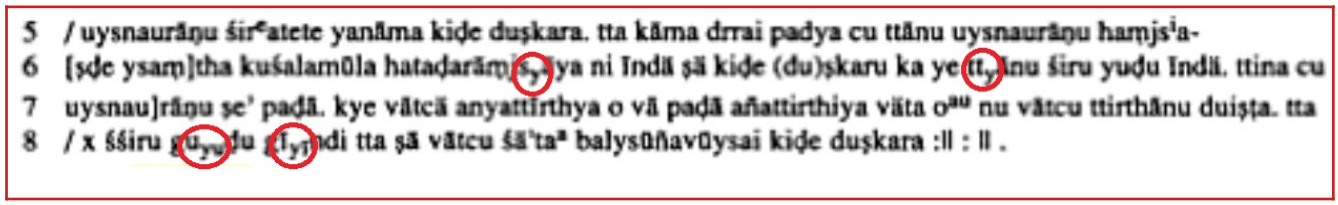


Figure 19. Skjærvø (2002: 343)

Subscript z (ᵢz)

Subscripts are used for fricative release in some French sources, filling the role played by superscripts in English. An example is subscript s and z for the frication of [tˢ] and [dᶻ] (allophones of /t/ and /d/ in Quebequois French). Also expected for foreign loans in Javanese (below).

Patricia Keating used it during her tenure as president of the IPA.

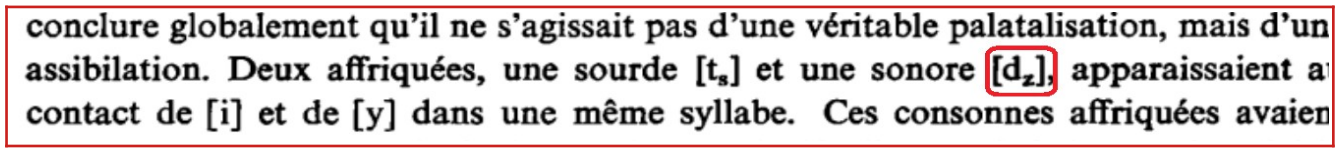


Figure 20. Charbonneau & Jacques (1972: 77). Subscript *z* is even used in the title of the article, "[tₛ] et [dᵢz] en français canadien."

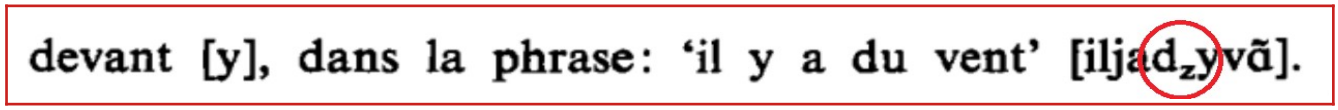


Figure 21. Charbonneau & Jacques (1972: 87)

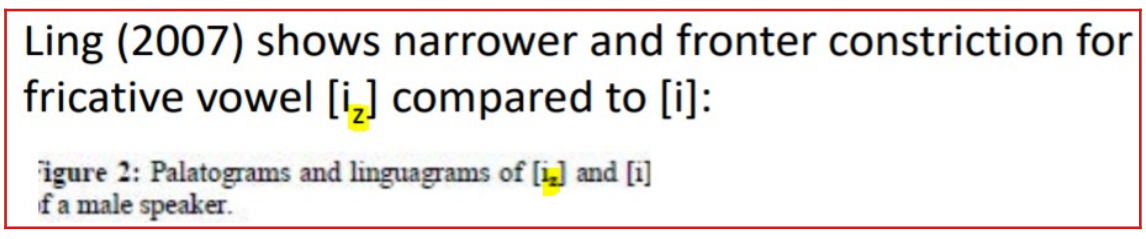


Figure 22. Keating (2018: 27).

There are totally 12 vowels in Suzhou Chinese, which are [i$_{z}$ y$_{z}$ u i y ø ɛ o æ ɑ ɿ ʯ].Two pairs of

vowel quality, the test words associated with high level tone [44] with zero initial consonant were selected. The test words were:
[i$_{z}^{44}$] (coat) [i^{44}] (smoke).

Figure 23. Ling (2007: 573)

Subscript ɒ and ʊ (ₒ ᵤ)

#$_{d}$ $_{\#}$d$_{ɒ}$ $_{d}$ɒ$_{g}$ $_{ɒ}$g$_{z}$ $_{g}$z$_{\#}$ $_{z}$\#$_{m}$ $_{\#}$m$_{ei}$ $_{m}$ei$_{k}$ $_{ei}$k$_{\#}$ $_{k}$\#$_{g}$ $_{\#}$g$_{ʊ}$ $_{g}$ʊ$_{d}$ $_{ʊ}$d$_{\#}$ $_{d}$\#p $_{\#}$p$_{e}$ $_{p}$e$_{t}$ $_{e}$t$_{s}$ $_{t}$s$_{\#}$ $_{s}$\#

Figure 24. Tatham & Morton (2016: 59). Subscript # and *ɒ d g ʊ z* for English.

Subscript ʃ (ᵪ)

Wickelgren could only index his objects according to their immediate linear context: $_{\#}$χ$_{o}$ χ$_{o}$r$_{o}$ r$_{o}$ ɾ$_{o}$ʃ$_{o}$ʃ$_{o}$#. This will not do because the variants must be derived *before* the representation not after it.

Figure 25. Tatham & Morton (2016: 60). Subscript # and ʃ for Russian.

Subscript ð ɛ f ʊ ʌ (ₔ ₑ f ᵤ ᵥ)

unordered set for the word /strʌk/ is not identical to that for the word /krʌst/: the former set is /$_{\#}$s$_{t}$, $_{s}$t$_{r}$, $_{t}$r$_{ʌ}$, $_{r}$ʌ$_{k}$, $_{ʌ}$k$_{\#}$/, whereas the latter set is /$_{\#}$k$_{r}$, $_{k}$r$_{ʌ}$, $_{r}$ʌ$_{s}$, $_{ʌ}$s$_{t}$, $_{s}$t$_{\#}$/. Hence, the partial activation of the unordered set of context-sensitive emr's is sufficient

Figure 26. Lass (2012: 86). Subscript # and ʌ for English.

ample, in the phrase, "the fretful elk," which we shall now represent allophonically as /$_{\#}$ð$_{ʌ}$, ðʌ$_{f}$, $_{ʌ}$f$_{r}$, $_{f}$r$_{ɛ}$, $_{r}$ɛ$_{t}$, ɛt$_{f}$, $_{t}$f$_{u}$, $_{f}$U$_{l}$, ʊl$_{ɛ}$, $_{l}$ɛ$_{l}$, ɛl$_{k}$, $_{l}$k$_{\#}$/, the word representatives partially activate almost this entire unordered set. However, instead of the specific allophone representatives, $_{\#}$ð$_{ʌ}$, ðʌ$_{f}$, $_{ʌ}$f$_{r}$, ʊlɛ, $_{l}$ɛ$_{l}$, and $_{l}$k$_{\#}$, the word representatives even less

Figure 27. Wickelgren (1969: 14). Subscript # and *ð ɛ f ʊ ʌ* for English. Subscript *ð* is overly large, but that is merely a typesetting issue.

Subscript Latin phi (ᵩ)

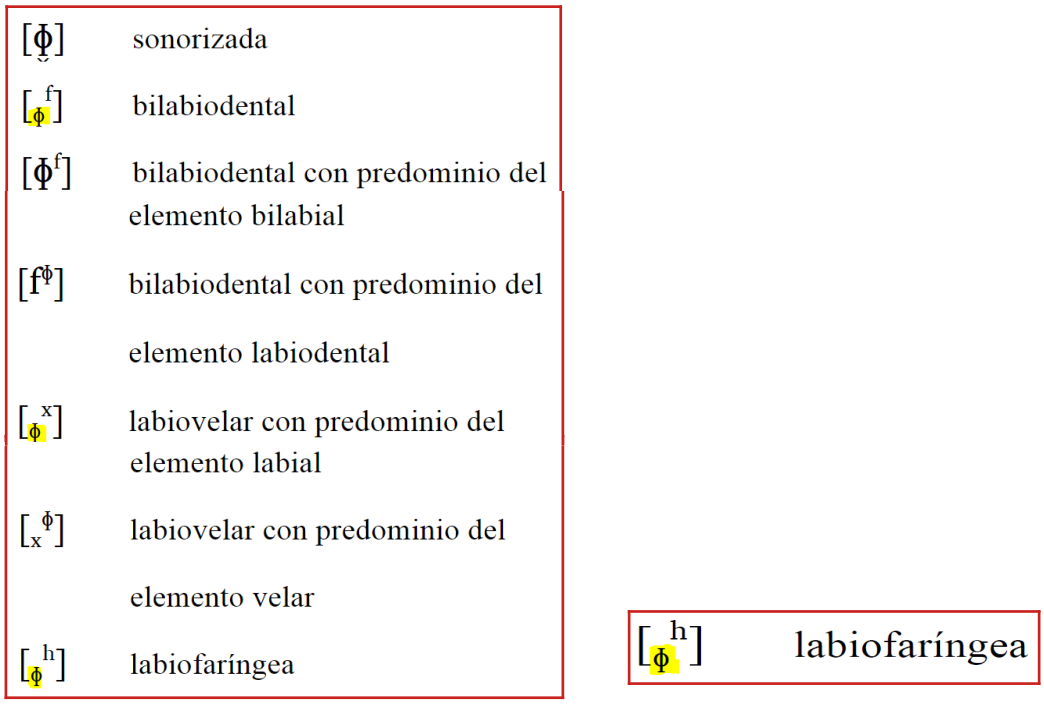


Figure 28. Butragueño p. 29 *ff.* IPA equivalents of the RFE Spanish phonetic alphabet, as used e.g. in the *Atlas Lingüístico de México*, contrasting baseline, superscript and subscript ɸ.

Subscript mu (ᵤ)

〈μ〉 is a wild card for 'mora', whereas subscript 〈ᵤ〉 marks a segment as moraic. The distinction is analogous to 〈σ〉 for 'syllable' vs the diacritic 〈◌̩〉 for syllabic.

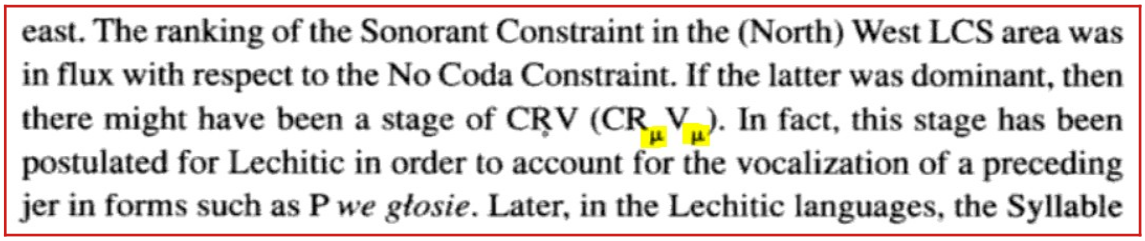

east. The ranking of the Sonorant Constraint in the (North) West LCS area was in flux with respect to the No Coda Constraint. If the latter was dominant, then there might have been a stage of CR̩V (CR$_{\mu}$V$_{\mu}$). In fact, this stage has been postulated for Lechitic in order to account for the vocalization of a preceding jer in forms such as P *we głosie*. Later, in the Lechitic languages, the Syllable

Figure 29. Bethin (1998: 250). The subscript μ's clarify that both the liquid R and the vowel V are moraic in the 〈CR̩V〉 syllable.

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*
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: *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): *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-G, Greek and Coptic*
2. Number of characters in proposal: *23*
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 or Michael Everson*
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):
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 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? *proposer is a member of 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? *publishing*
   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:
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)? *preferably*
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: