L2/21-042

Unicode request for phonetic punctuation & diacritics

Kirk Miller, kirkmiller@gmail.com 2021 January 11

This proposal is for various non-IPA conventions used in the phonetic and phonological literature.

Thanks to Deborah Anderson of the Universal Scripts Project for her assistance.

Combining diacritics

◌᪵ 1AC5 COMBINING SQUARE BRACKETS ABOVE. Figures 13, 14, 16.

◌᪷ 1AC7 COMBINING INVERTED DOUBLE ARCH ABOVE. Figures 1–4.
Cf. 032B COMBINING INVERTED DOUBLE ARCH BELOW,
cf. 1DF1 COMBINING LATIN SMALL LETTER W,
cf. 1ABF COMBINING LATIN SMALL LETTER W BELOW

◌᪸ 1AC8 COMBINING PLUS SIGN ABOVE. Figures 5–8.
Cf. 031F COMBINING PLUS SIGN BELOW.

◌᪽ 1ACD COMBINING DOUBLE PLUS SIGN ABOVE. Figure 11.

◌᪾ 1ACE COMBINING DOUBLE PLUS SIGN BELOW. Figures 9–10.
Cf. 031F COMBINING PLUS SIGN BELOW.

Punctuation marks

⸢ 2E55 LEFT SQUARE BRACKET WITH STROKE. Figures 12, 14.
Cf. 2045 LEFT SQUARE BRACKET WITH QUILL ⁅

⸣ 2E56 RIGHT SQUARE BRACKET WITH STROKE. Figures 12, 14.
Cf. 2046 RIGHT SQUARE BRACKET WITH QUILL ⁆

⸤ 2E57 LEFT SQUARE BRACKET WITH DOUBLE STROKE. Figures 12, 15.

⸥ 2E58 RIGHT SQUARE BRACKET WITH DOUBLE STROKE. Figures 12, 15.

⸜ 2E59 TOP HALF LEFT PARENTHESIS. Figure 17.

⸝ 2E5A TOP HALF RIGHT PARENTHESIS. Figure 17.

⸞ 2E5B BOTTOM HALF LEFT PARENTHESIS. Figure 17.

⸟ 2E5C BOTTOM HALF RIGHT PARENTHESIS. Figures 17–18.

Spacing diacritics

ʰ 11AB0 MODIFIER NUMBER SIGN. Figures 19–24.

ʲ 11AB1 MODIFIER DOLLAR SIGN. Figure 25–26.

Encoding order

Because the combining brackets U+1AC5 ◌᷅ are intended to modify another diacritic, their behaviour should be the same as U+1ABB ◌᪻ COMBINING PARENTHESES ABOVE, as described in the section *Combining Diacritical Marks Extended: U+1AB0–U+1AFF* of TUS, on p. 331 and figure 7-13. IPA usage should be added to the text in TUS.

Properties

```
1AC5;COMBINING SQUARE BRACKETS ABOVE;Mn;230;NSM;;;;;N;;;;;
1AC7;COMBINING INVERTED DOUBLE ARCH ABOVE;Mn;230;NSM;;;;;N;;;;;
1AC8;COMBINING PLUS SIGN ABOVE;Mn;230;NSM;;;;;N;;;;;
1ACD;COMBINING DOUBLE PLUS SIGN ABOVE;Mn;230;NSM;;;;;N;;;;;
1ACE;COMBINING DOUBLE PLUS SIGN BELOW;Mn;220;NSM;;;;;N;;;;;
2E55;LEFT SQUARE BRACKET WITH STROKE;Ps;0;ON;;;;;Y;;;;;
2E56;RIGHT SQUARE BRACKET WITH STROKE;Pe;0;ON;;;;;Y;;;;;
2E57;LEFT SQUARE BRACKET WITH DOUBLE STROKE;Ps;0;ON;;;;;Y;;;;;
2E58;RIGHT SQUARE BRACKET WITH DOUBLE STROKE;Pe;0;ON;;;;;Y;;;;;
2E59;TOP HALF LEFT PARENTHESIS;Ps;0;ON;;;;;Y;;;;;
2E5A;TOP HALF RIGHT PARENTHESIS;Pe;0;ON;;;;;Y;;;;;
2E5B;BOTTOM HALF LEFT PARENTHESIS;Ps;0;ON;;;;;Y;;;;;
2E5C;BOTTOM HALF RIGHT PARENTHESIS;Pe;0;ON;;;;;Y;;;;;
11AB0;MODIFIER NUMBER SIGN;Lm;0;L;<super> 0023;;;;N;;;;;
11AB1;MODIFIER DOLLAR SIGN;Lm;0;L;<super> 0024;;;;N;;;;;
```

Bidi values

The eight punctuation marks U+2E55 to U+2E5C have the bidi-mirrored property “Yes”. The following are the bidi-mirroring glyph values for BidiMirroring.txt:

```
2E55; 2E56 # LEFT SQUARE BRACKET WITH STROKE
2E56; 2E55 # RIGHT SQUARE BRACKET WITH STROKE
2E57; 2E58 # LEFT SQUARE BRACKET WITH DOUBLE STROKE
2E58; 2E57 # RIGHT SQUARE BRACKET WITH DOUBLE STROKE
2E59; 2E5A # TOP HALF LEFT PARENTHESIS
2E5A; 2E59 # TOP HALF RIGHT PARENTHESIS
2E5B; 2E5C # BOTTOM HALF LEFT PARENTHESIS
2E5C; 2E5B # BOTTOM HALF RIGHT PARENTHESIS
```

Chart

Characters in white cells are proposed here. Characters on light grey backgrounds have been approved by the UTC for Unicode 14.

	...0	...1	...2	...3	...4	...5	...6	...7	...8	...9	...A	...B	...C	...D	...E	...F
Combining Diacritical Marks Extended																
U+1ACx																
Supplemental Punctuation																
U+2E5x						⸢	⸣	⸤	⸥	⸜	⸝	⸞	⸟			
Spacing Modifier Letters-A																
U+11ABx	#	$														

References

Allen & Hawkins (1978) Development of Phonological Rhythm, in Bell & Hooper (eds.) *Syllables and Segments*.

Carlson (1972) *A Grammar of Spokan*.

Dolgopolsky (1987) South Cushitic Lateral Consonants as compared to Semitic and East Cushitic. *Proceedings of the Fourth International Hamito-Semitic Congress, Marburg, 20–22 September, 1983*.

Dolgopolsky (2013) *Indo-European Dictionary with Nostratic Etymologies*, vol. I.

Ellis (1889) *On Early English Pronunciation*, part V.

Fukui (2004) TIPA Manual, version 1.3.

Kretzschmar (1993/1994) *Handbook of the Linguistic Atlas of the Middle and South Atlantic States*. University of Chicago Press. [LAMSAS is a project of the University of Georgia]

Kelly & Local (1989) *Doing Phonology.* Manchester University Press.

Samuel Martin (2004) *A Reference Grammar of Japanese*, University of Hawaii Press.

McDavid & O'Cain (1980) *Linguistic Atlas of the Middle and South Atlantic States*. University of Chicago Press, fasc. 2.

Alexis Michaud (2008) 'Phonemic and tonal analysis of Yongning Na', *Cahiers de linguistique – Asie Orientale*, CRLAO 37 (2), pp. 159-196.

Alexis Michaud (2012) 'Monosyllabicization: patterns of evolution in Asian languages', in Stolz, Nau & Stroh, ed. *Monosyllables – from Phonology to Typology*, De Gruyter.

Penhallurick (1991) *The Anglo-Welsh Dialects of North Wales*.

Smith-Stark (2005) "Phonological description in New Spain", in Zwartjes & Altman (eds.) *Missionary Linguistics II (Lingüística misionera II): Orthography and Phonology*, John Benjamins.

Figures

Combining inverted double arch above (◌͝)

As with several other subscript IPA diacritics, ◌͝ (subscript script *w* ≈ subscript omega) may be placed above a letter with a descender such as *g* or *ŋ*: ⟨g̈ ŋ̈ χ̈⟩. This retired IPA diacritic remains useful for those who wish to distinguish simultaneous labialization from labialized release or, like Ladefoged & Maddieson, simple labialization from labio-velarization, as well as for typesetting older sources. Combining superscript script *w* / omega was also used in colonial missionary sources of Otomi, and may be useful in reproducing those materials. (I have not found any other unsupported symbols in that material: the Parra letters *tresillo*, *quatrillo* and *quatrillo*-comma and the various barred letters are all supported by Unicode.)

3.2.1 *Nasalization*. Many Mexican languages, above all those in the Otomanguean family, distinguish between oral and nasalized vowels. In the case of Otomi, as registered by the Franciscan friars Pedro de Cárceres (fl. 1580) and Alonso Urbano, nasalized vowels were distinguished by writing a small omega-shaped diacritic above them, ⟨◌̽⟩, which, following a suggestion by Heriberto Avelino (personal communcation), might be called a *little bat*, or 'murcielaguito' in Spanish.

Figure 1. Smith-Stark (2005: 19)

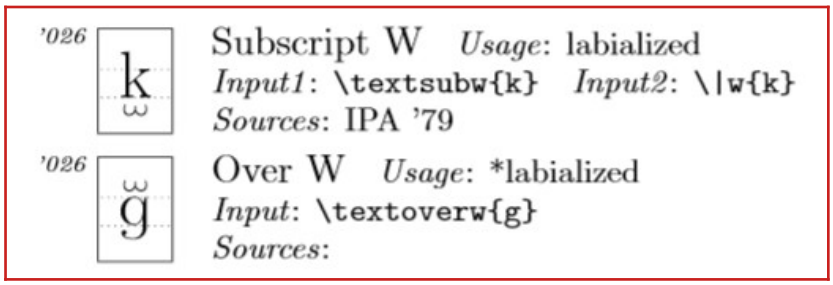


Figure 2. Fukui (2004: 52)

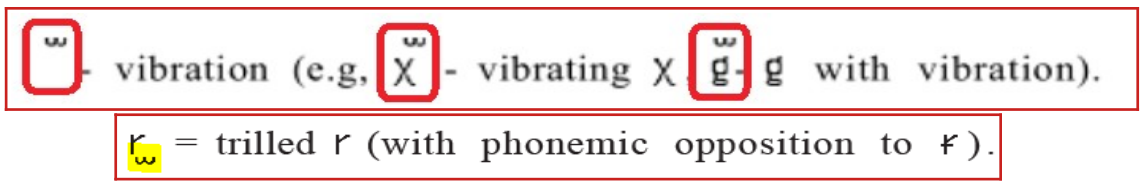


Figure 3. Dolgopolsky (2013: 239). Although semantically different from old IPA use, the diacritic is still only made superscript with letters such as *χ* and *g* that have a descender, unlike e.g. the alveolar trill ⟨r̝⟩. (Though slightly misaligned here on the *χ*, the diacritic is not intended to be used as a spacing diacritic after the letter. That would indicate labialization, as shown further up on p. 239.)

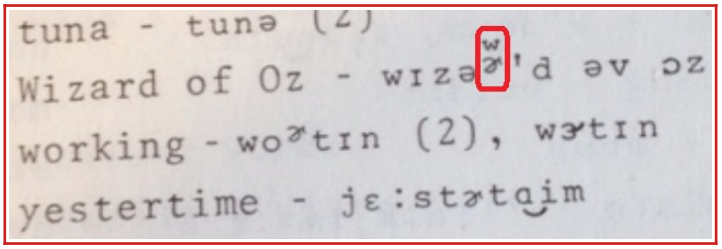


Figure 4. [wɪzəˠʷˈd əv ɔz] in Allen & Hawkins (1978: 183). The labialization diacritic is placed over the superscript ˠ, presumably to avoid confusion as to which is the base letter: on p. 184 they have the transcription ['o:vˠ̟] with the diacritic under the same letter where that is not superscript. The authors use a print *w* for the diacritic in the appendix (p. 182–185), but in the text they write it in by hand as a script *w* that more closely resembles an omega.

Combining plus sign above (◌̟)

An allograph of the IPA subscript diacritic, used to avoid descenders or other diacritics.

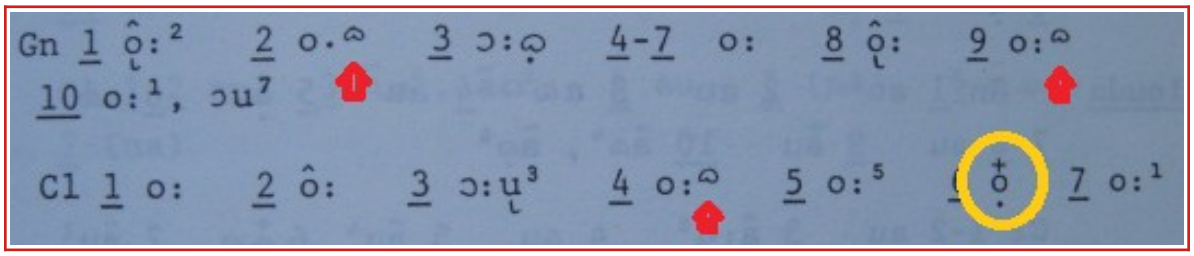


Figure 5. Penhallurick (1991: 79). Yellow ring ⟨o̟⟩. The advanced diacritic placed above the 'o' avoids conflict with the underdot.

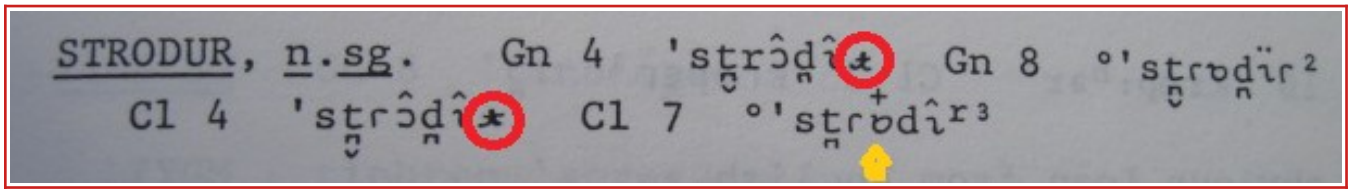


Figure 6. Penhallurick (1991: 251), yellow arrow marks ⟨o̟⟩. The '+' is consistently superscripted in this doc.

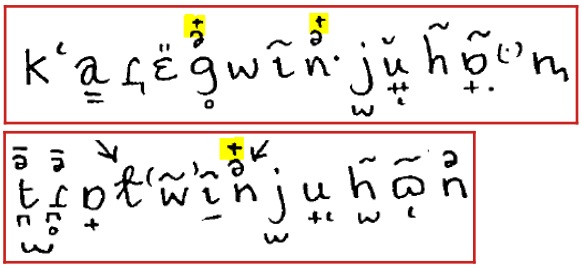


Figure 7. Kelly & Local (1989: 37). Here the plus diacritic is superscript because there is no room for it below the combining schwa. It contrasts with a superscript minus.

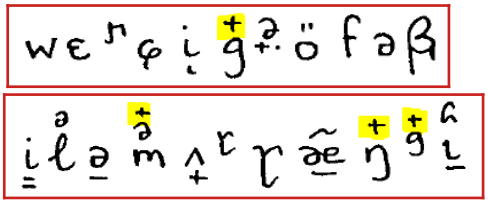


Figure 8. Kelly & Local (1989: 67, 70). Superscript plus to avoid clashing with a descender.

Combining double plus sign diacritics (◌̤, ◌̪)

Side-by-side doubling such as this requires specific Unicode support, as has already been provided for the double ‘open’ diacritic seen in the following figure.

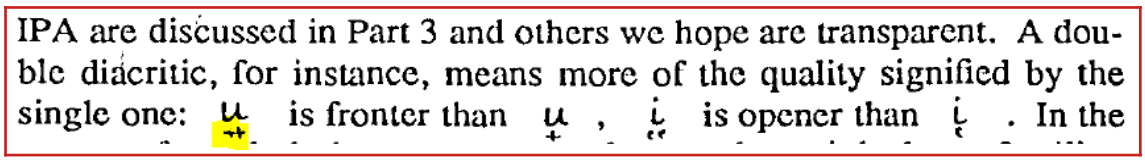


Figure 9. Kelly & Local (1989: 8). Although in handwriting two plus signs may join together, logically they are double, like double ‘open’ here and double minus next. (Double ‘open’ is supported at U+1AB8. See Fig. 5 for a typeset single ‘open’ diacritic.)

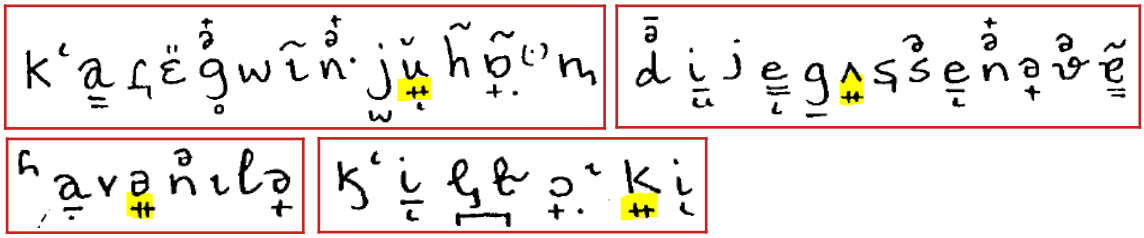


Figure 10. Kelly & Local (1989: 37, 70, 72, 88). Note the stacking of diacritics with the double-plus on the ⟨u̪⟩.

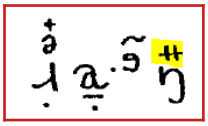
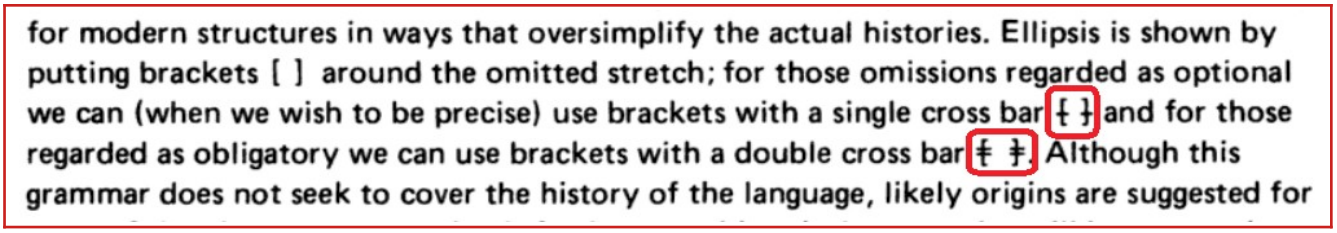


Figure 11. Kelly & Local (1989: 156). As with the single plus sign, the double plus appears over a letter with a descender.

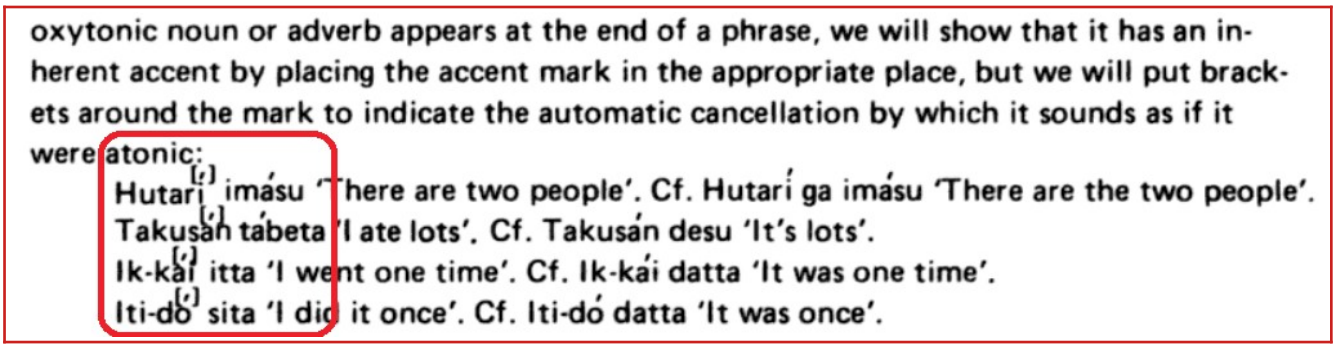
Barred square brackets (⟦...⟧ ⧼...⧽) and combining square brackets above (͡)

Used in Martin (2004) for ellipsis in Japanese. ͡ is only attested as a pair, so a single code point might be preferred.



for modern structures in ways that oversimplify the actual histories. Ellipsis is shown by putting brackets [ ] around the omitted stretch; for those omissions regarded as optional we can (when we wish to be precise) use brackets with a single cross bar ⟦ ⟧ and for those regarded as obligatory we can use brackets with a double cross bar ⧼ ⧽. Although this grammar does not seek to cover the history of the language, likely origins are suggested for

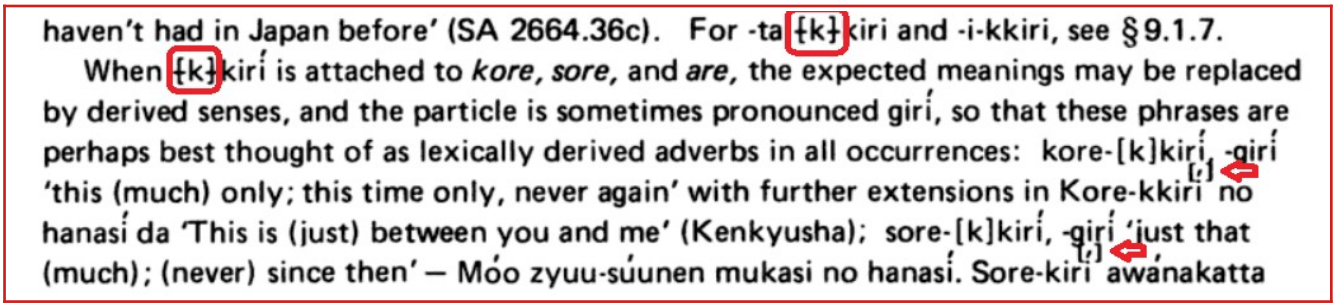
Figure 12. Samuel Martin (2004: 28)



oxytonic noun or adverb appears at the end of a phrase, we will show that it has an inherent accent by placing the accent mark in the appropriate place, but we will put brackets around the mark to indicate the automatic cancellation by which it sounds as if it were atonic:

Hutari[ˊ] imásu 'There are two people'. Cf. Hutarí ga imásu 'There are the two people'.
Takusán[ˊ] tábeta 'I ate lots'. Cf. Takusán desu 'It's lots'.
Ik-kái[ˊ] itta 'I went one time'. Cf. Ik-kái datta 'It was one time'.
Iti-dó[ˊ] sita 'I did it once'. Cf. Iti-dó datta 'It was once'.

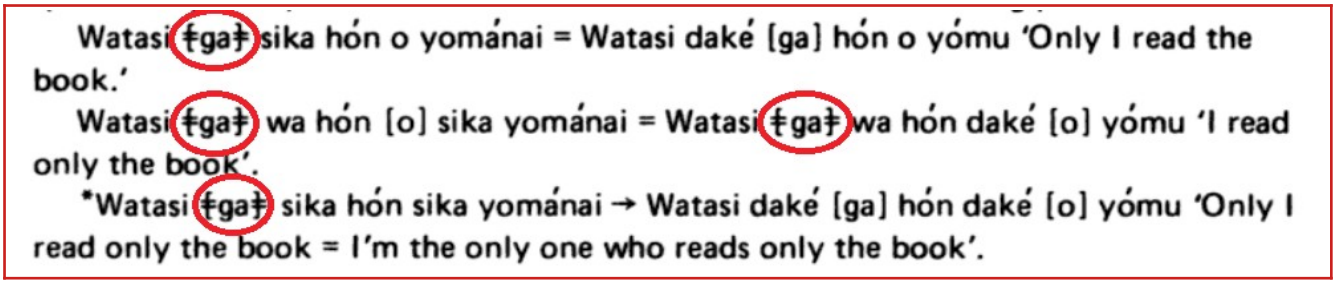
Figure 13. Martin (2004: 21).



haven't had in Japan before' (SA 2664.36c). For -ta ⟦k⟧kiri and -i-kkiri, see §9.1.7.

When ⟦k⟧kirí is attached to *kore*, *sore*, and *are*, the expected meanings may be replaced by derived senses, and the particle is sometimes pronounced girí, so that these phrases are perhaps best thought of as lexically derived adverbs in all occurrences: kore-[k]kirí[ˊ], -girí 'this (much) only; this time only, never again' with further extensions in Kore-kkiri no hanasí da 'This is (just) between you and me' (Kenkyusha); sore-[k]kirí[ˊ], -girí 'just that (much); (never) since then' — Móo zyuu-súunen mukasi no hanasí. Sore-kirí awánakatta

Figure 14. Martin (2004: 78)

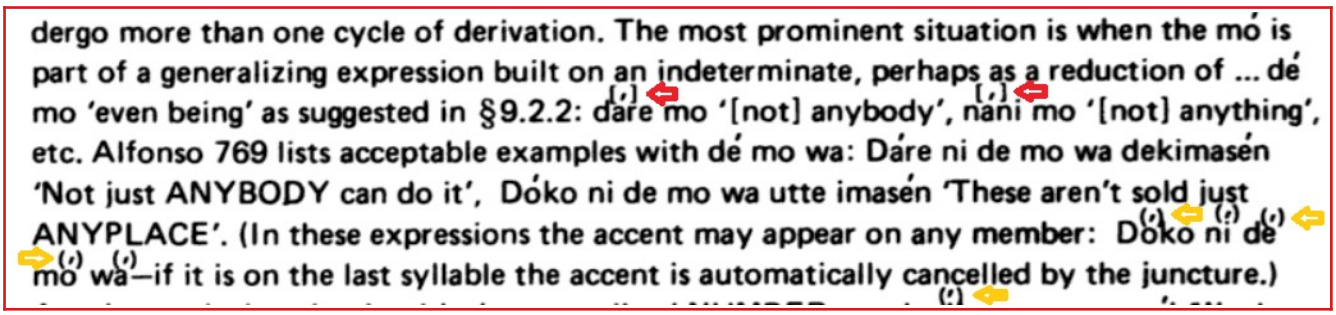


Watasi ⟦ga⟧ sika hón o yománai = Watasi daké [ga] hón o yómu 'Only I read the book.'

Watasi ⟦ga⟧ wa hón [o] sika yománai = Watasi ⟦ga⟧ wa hón daké [o] yómu 'I read only the book'.

*Watasi ⟦ga⟧ sika hón sika yománai → Watasi daké [ga] hón daké [o] yómu 'Only I read only the book = I'm the only one who reads only the book'.

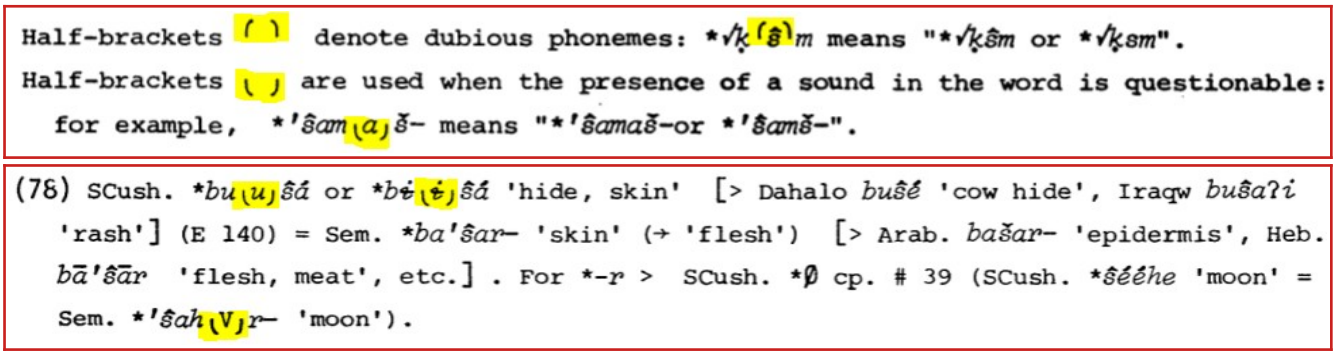
Figure 15. Martin (2004: 77)



dergo more than one cycle of derivation. The most prominent situation is when the mó is part of a generalizing expression built on an indeterminate, perhaps as a reduction of ... dé mo 'even being' as suggested in §9.2.2: dáre mo '[not] anybody', nání mo '[not] anything', etc. Alfonso 769 lists acceptable examples with dé mo wa: Dáre ni de mo wa dekimasén 'Not just ANYBODY can do it', Dóko ni de mo wa utte imasén 'These aren't sold just ANYPLACE'. (In these expressions the accent may appear on any member: Dóko ni dé mo wa—if it is on the last syllable the accent is automatically cancelled by the juncture.)

Figure 16. Martin (2004: 54). Square brackets (red) are used for ellipsis of fixed stress; parentheses (yellow) for ellipsis of mobile stress.

Half parentheses (ˊ...ˎ ˏ...ˊ)

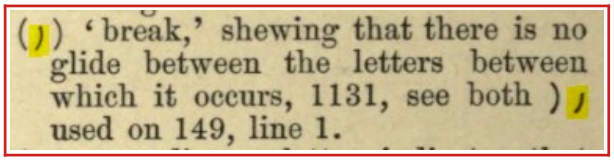


Half-brackets ˊ ˎ denote dubious phonemes: *√ḳˊŝˎm means "*√ḳŝm or *√ḳsm".
Half-brackets ˏ ˊ are used when the presence of a sound in the word is questionable: for example, *'ŝamˏaˊš- means "*'ŝamaš-or *'ŝamš-".

(78) SCush. *buˏuˊŝá or *bɨˏɨˊŝá 'hide, skin' [> Dahalo *buŝé* 'cow hide', Iraqw *buŝa?i* 'rash'] (E 140) = Sem. *ba'ŝar- 'skin' (→ 'flesh') [> Arab. *bašar-* 'epidermis', Heb. *bā'ŝār* 'flesh, meat', etc.] . For *-r > SCush. *Ø cp. # 39 (SCush. *ŝéé́he* 'moon' = Sem. *'ŝahˏvˊr- 'moon').

'heart', Galla *'lubb-a?* 'life', Konso *lup'p-oota* 'heart, soul'] (B 259) = Central Cush. *lvˊbbˋʌk- 'heart' [> Bilin (P) *lʌbbʌkʌ*, pl. *lʌffʌk* 'heart, mind', Kwara (R) *lʌbakā*, Kemant (CR) *ləbʌkā*, (Bn) *ləbʌʌka* 'heart'] (DSF 163-4).

Figure 17. Dolgopolsy (1987: 212, 205, 198)



(ˊ) 'break,' shewing that there is no glide between the letters between which it occurs, 1131, see both) ˊ used on 149, line 1.

Figure 18. Ellis (1889: V.88*).

Modifier number sign (#)

The words for 'calf', 'mare' and 'stallion' in Yongning Na illustrate different stages along the path towards full lexicalization. (About the notation of tones in the following examples, see section 4 of Michaud 2008.) 'Calf', /#Hʐwæ.zo/, literally means 'baby horse': it is made up of 'horse', /#Hʐwæ/, plus the word for 'son', /#Hzo/; the latter is clearly on its

Figure 19. Michaud (2012: 127). Superscript #H is a boundary tone (a floating high tone, which can only exist after #, a word boundary), not just a high tone H that happens to appear after a word boundary, which is what currently supported [#H] would indicate.
[The apparent example of 〈#〉 in Figure 2 was intended to be an overscript.]

	Yongning Na	Western Naxi
1. 'water, shuǐ 水'	/Ldʑi/	/Lgi/
2. 'to fall (rain), 下 (雨, 雪) xià (yǔ, xuě)'	/gi/ (*toneless*[5])	/Mgɯ/
3. 'bear (animal), 熊 xióng'	/#Hgi.nɑ.mi/	/Lgɣ/
4. 'granary, 粮仓 liángcāng'	/Lgi /	/Lŋgu/
5. 'little brother, 弟弟 dìdi'	/#Hgi.zɯ/	/Mgɯ Mzɯ/
6. 'half, (一) 半 (yí) bàn'	/Hgi/	/Mŋgɯ/
7. existential/'to have, 有 yǒu'	/Mdʑu/	/Mgy/

Figure 20. Michaud (2008: 7). Phonemic contrast between /H/ and /#H/.

tone of the root with the entire disyllable (*ibid.*, 192); in the case of 'mare', one would expect a #H tone, not a L tone. Lastly, 'stallion', /MLʐwæ.sɯ/, is yet further advanced towards independent existence as a

Figure 21. Michaud (2008: 25).

Thus, [ʁuˈdzɯ˩] 'Tibetan' can be analysed as /#Lʁu.dzɯ/ or /MLʁu.dzɯ/. At present, no decisive evidence has been found in favour of one analysis over the other. An argument against the contour interpretation is that no [ML] contour ever appears on a monosyllable, unlike [LM] and [MH]. Decisive arguments in favour

Figure 22. Michaud (2008: 32). Theoretical distinction between /#L/ and /ML/, though both are phonetically [ML].

Table 9. The tones of monosyllabic nouns in Na.

tonal analysis	example	
M	^{M} lɑ	‘tiger’
L	^{L} kʰɣ̩	‘dog’
LM	^{LM} bu	‘pig’
MH	^{MH} hwɤ	‘cat’ [vocative only]
#H	$^{\#H}$ ʐ̩wæ	‘horse’

Figure 23. Michaud (2008: 26). /$^{\#H}$/ indicates a word has a floating high tone. When tones are indicated for themselves rather than prefixed to a particular a word, they’re baseline L, M, H, #H (left column).

enough to allow for a loan-translation into the Na language: to their native word /^{LM}nɑ/, they add their own word for ‘man’, /$^{\#H}$ hĩ/, yielding /$^{LM+\#H}$nɑ.hĩ/, a disyllable with a complex tonal pattern, /LM+#H/, *i.e.* a LM contour plus a floating H tone.

Figure 24. Michaud (2008: 31). Explanation of the symbol.

Modifier dollar sign ($)

[L.LM]	L	**L.L+M**	**L.L+M**	/^{L} kʰɣ̩.mi/	[kʰɣ̩˩ mi˧˥]	‘she-dog, bitch’
[M.M]	#H	**M.M+M**	**M.M+H**	/$^{\#H}$ ŋi.mi/	[ŋi˧ mi˧]	‘sun’
[M.MH]	MH#	**M.M+H**	**M.M+H**	/$^{MH\#}$ hwɤ.li/	[hwɤ˧ li˥]	‘cat’
[L.M]	LM	**L.M+M**	**L.M+L**	/^{LM} bu.mi/	[bu˩ mi˧]	‘sow, female pig’
[M.H]	H$	**M.M+M**	**M.M+H**	/$^{H\$}$ hwɤ.mi/	[hwɤ˧ mi˥]	‘she-cat’
[L.M]	LM+#H	**L.M+M**	**L.M+H**	/$^{LM+\#H}$ nɑ.hĩ/	[nɑ˩ hĩ˧]	‘Naxi’
[L.MH]	L+MH	**L.M+H**	**L.M+H**	/$^{L+MH}$ i.ʈʂæ/	[i˩ ʈʂæ˥]	‘waist’

Figure 25. Michaud (2008: 29). Contrast between /$^{H\$}$/, a high tone characteristic of the end of a phonological phrase, and /$^{H\#}$/, a high tone characteristic of the end of a word.

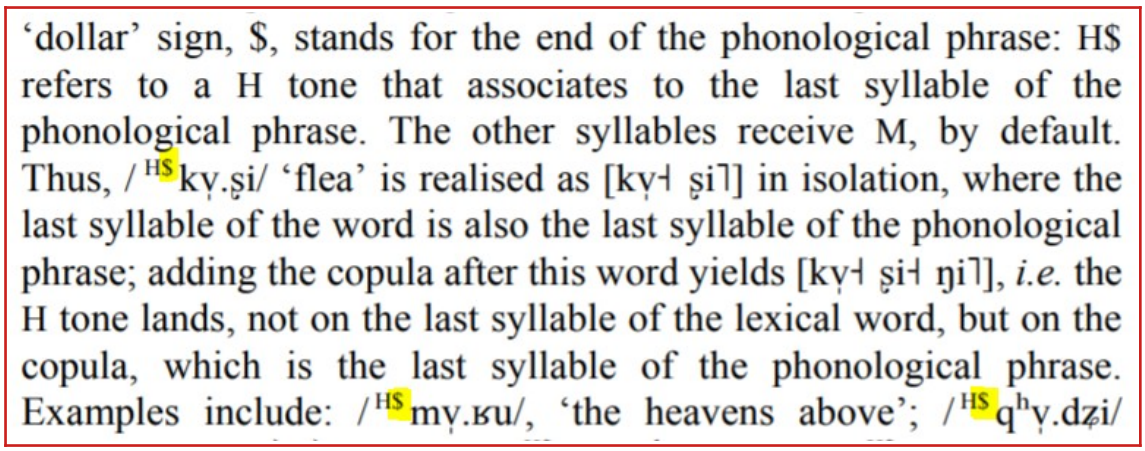

‘dollar’ sign, $, stands for the end of the phonological phrase: H$ refers to a H tone that associates to the last syllable of the phonological phrase. The other syllables receive M, by default. Thus, /ᴴˢkɣ̣.ʂi/ ‘flea’ is realised as [kɣ̣˦ ʂi˥] in isolation, where the last syllable of the word is also the last syllable of the phonological phrase; adding the copula after this word yields [kɣ̣˦ ʂi˦ ŋi˥], *i.e.* the H tone lands, not on the last syllable of the lexical word, but on the copula, which is the last syllable of the phonological phrase. Examples include: /ᴴˢmɣ̣.ʁu/, ‘the heavens above’; /ᴴˢqʰɣ̣.dʑi/

Figure 26. Michaud (2008: 31). Explanation of the symbol.

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 IPA letters (a), pulmonic*
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):
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *some*
      Name of the existing block: *Combining Diacritical Marks Extended, Supplemental Punctuation, Spacing Modifier Letters-A (new)*
2. Number of characters in proposal: *15*
3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary √ 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)? *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.)? *no*
   If YES, with whom?
   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: *see References section*
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? *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: