

# Unicode request for Harrington diacritic

Kirk Miller, kirkmiller@gmail.com

Timothy Henry-Rodriguez, thenry@fullerton.edu

2020 July 11

For many languages, Harrington’s materials are the most extensive and reliable documentation available. Increasing use of his material is being made by indigenous communities for language-revival projects, such as a Purismeno Chumash dictionary that one of us (Henry-Rodriguez) has worked on and that is currently under review by the community. However, lack of font support has meant that this diacritic has been rendered by a simple octothorp. The community has debated leaving it out of the dictionary altogether for fear that including it as a spacing diacritic would misrepresent the material more than leaving it out. Figure 2 illustrates the problem.

The Harrington Database Project at UC Davis (Martha Macri, Victor Golla, Justin Spence, Lisa Woodward) transcribe Harrington materials using a makeshift encoding system in the PUA. They have been contemplating a Unicode submission, but say (p.c.) that the identities of many of the symbols are still doubtful, and it would be premature to add most Harrington symbols to Unicode. However, project members agree that the identity of the superscript octothorp is secure and it would be good to add it to Unicode since it is needed beyond the database project.

## Combining diacritic

# U+1AC6 COMBINING NUMBER SIGN. Figures 1–2 and letter from the Western Institute for Endangered Language Documentation as page 3.

## Properties

1AC6;COMBINING NUMBER SIGN;Mn;230;NSM;;;;;N;;;;;

## Chart

The highlighted character is proposed in this request. Greyed-out cells are already assigned. Other characters are being proposed in separate requests by one of the authors.

	...0	...1	...2	...3	...4	...5	...6	...7	...8	...9	...A	...B	...C	...D	...E	...F
Combining Diacritical Marks Extended																
U+1ACx		◌̆	◌̇	◌̈	◌̉	◌̊	◌̋	◌̌	◌̍							

## Reference

Suzanne Wash (2001) *Adverbial Clauses in Barbareño Chumash Narrative Discourse*. PhD dissertation, University of California Santa Barbara.

## Figures

### Combining number sign (◌<sup>#</sup>)

Used by Harrington for strong stress, and by linguists and archivists transcribing his work.

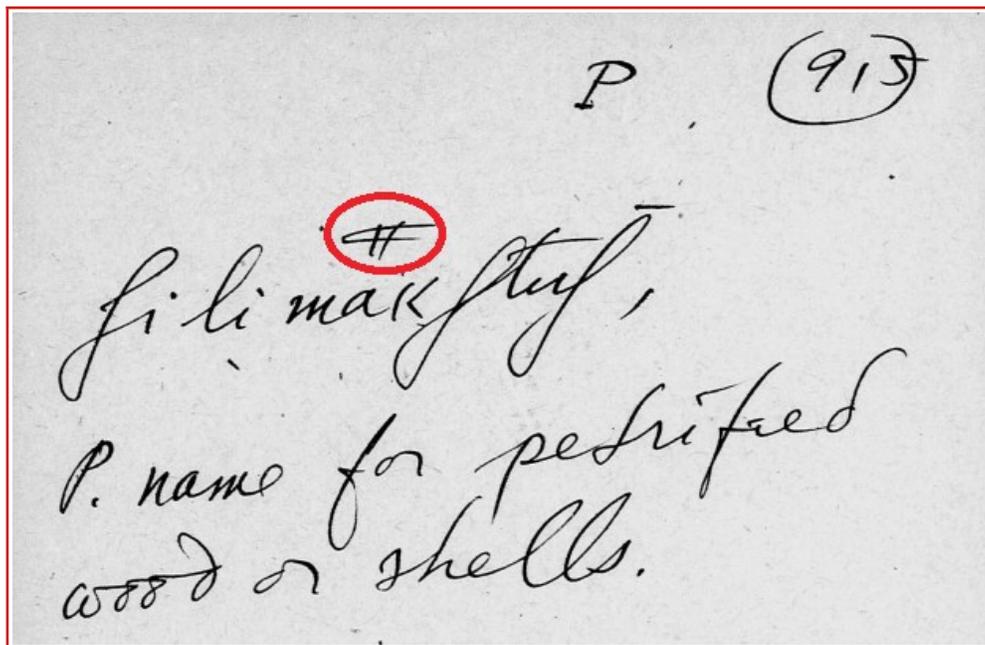


Figure 1. Harrington index card for Purisimeño Chumash showing the proper placement of the diacritic on the word *filimakstuf* (<filim<sup>#</sup>akstuf>).

- (134) **pa<sup>#</sup>** ʔalxintší·wa`š hisxá·la`š. hihoʔsmət  
pa ʔal-xinči-waš hi=s-xalas hi=hoʔ=s-mit  
EVID STV-bad -PST DP=3-heal DP=DIS=3-back  
'It evidently healed wrongly'  
<Evidently her back healed badly,> 59.0647
- (135) ʔasanneswäš **ke<sup>#</sup>** his.amekwəl hilstük.  
ʔa=s-am-nes-waš ke hi=s-am -eqwel hi=l =stük  
Q =3-IDF-do-PST Q DP=3-IDF-make DP=ART=bowl.type  
'como hacian un s.tuk.'  
<How did they make a stuk?> 59.0368, elicitation

Figure 2. Wash (2001: 47). The diacritic # should appear over the preceding letter in <p<sup>#</sup>a> and <k<sup>#</sup>a>, but lack of font support prevented proper formatting. Similarly, <wa`š> and <la`s> are approximations of <wäš> and <làs>.



# Western Institute for Endangered Language Documentation

<http://www.wieldoc.org>

7414 Magnolia Ave, #3  
Riverside, CA 92504  
USA

July 1, 2020

Unicode Script Ad Hoc Committee

Greetings:

The Western Institute for Endangered Language Documentation supports Dr. Timothy Henry-Rodriguez's proposal to have a combined pound sign above added to the characters supported by Unicode. This is a common symbol found in Harrington material, and its use is widespread in marking primary stress and aspects of intonation in the languages Harrington documented. Dr. Henry-Rodriguez composed WIELD's initial lexicon of Purisimeño Chumash, and he is currently expanding that project for a National Endowment for the Humanities Documenting Endangered Languages grant. The importance of the accurate representation of original fieldwork notes is paramount.

WIELD also supports the proposal for the addition of a combining pound sign above as an intern in our corporation is compiling a series of texts (parsed) for Yowlumne (Yawelmani Yokuts), and, for Harrington's notes on this language, the combining pound sign above is also used as a marker of stress and intonation.

WIELD's board of directors has unanimously approved the proposal to add support for a combining pound sign above to the Unicode table. Please feel free to contact a representative of the board for further comment (contact information below).

Sincerely,

Uldis Balodis, PhD  
Board President  
[ubalodis@gmail.com](mailto:ubalodis@gmail.com)

Neil Alexander Walker, PhD  
Board Vice-President  
[neil.alexander.walker@gmail.com](mailto:neil.alexander.walker@gmail.com)

ISO/IEC JTC 1/SC 2/WG 2  
**PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
 FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>**

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**

<b>1. Title:</b>	<i>Additional phonetic click letters</i>	
<b>2. Requester's name:</b>	<i>Kirk Miller, Timothy Henry</i>	
<b>3. Requester type (Member body/Liaison/Individual contribution):</b>	<i>Western Institute for Endangered Language Documentation</i>	
<b>4. Submission date:</b>	<i>2020 July 11</i>	
<b>5. Requester's reference (if applicable):</b>		
<b>6. Choose one of the following:</b>		
This is a complete proposal:		<input checked="" type="checkbox"/>
(or) More information will be provided later:		<input type="checkbox"/>

**B. Technical - General**

<b>1. Choose one of the following:</b>			
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:		<input checked="" type="checkbox"/>	
Name of the existing block:	<i>Combining Diacritical Marks Extended</i>		
<b>2. Number of characters in proposal:</b>		<i>1</i>	
<b>3. Proposed category (select one from below - see section 2.2 of P&amp;P document):</b>			
A-Contemporary	<input 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"/>	B.2-Specialized (large collection)	<input checked="" type="checkbox"/>
<b>4. Is a repertoire including character names provided?</b>		<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>	
<b>5. Fonts related:</b>			
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>		
<b>6. References:</b>			
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>	
<b>7. Special encoding issues:</b>			
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>yes</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.

<sup>1</sup> 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? If YES explain	  	<i>no</i>
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.)? If YES, with whom? If YES, available relevant documents:	 <i>Harrington Database Project, WIELD</i> <i>WIELD letter attached, Harrington Project emails available on request</i>	<i>yes</i>
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) Reference:	  	<i>phonetic</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	 <i>see illustrations</i>	<i>yes</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	  	<i>no</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?		<i>no</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	  <i>(annotations to support those that are)</i>	<i>no</i>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	  <i>(Unicode disprefers use of combining retroflex tail)</i>	<i>yes</i>
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? If YES, is a rationale for its inclusion provided? If YES, reference:	  <i>(see comment by Script Ad Hoc Committee on integral sign)</i>	<i>yes</i> <i>yes</i>
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference:	  	<i>no</i>
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	  	<i>no</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	  	<i>no</i>
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	  	<i>no</i>