

Universal Multiple-Octet Coded Character Set  
International Organization for Standardization  
Organisation internationale de normalisation  
Международная организация по стандартизации

**Doc Type: Working Group Document****Title: Proposal to encode five Indo-Europeanist phonetic characters in the UCS****Source: Deborah Anderson and Michael Everson****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2004-02-02**

A number of subscripted letters are used in Indo-European linguistic materials, though their use is by no means limited to Indo-European. While a few Latin letters are currently encoded as modifier letters, the needs for Indo-European extend beyond this set. Five characters are proposed here. Styled text is not seen as appropriate for these; Indo-Europeanists already make use of the subscript digits, and superscript <sup>h</sup> and <sup>w</sup> and so on, already encoded. The characters proposed here are required for plain-text representation of Indo-European reconstructed material.

1D7A	a	LATIN SUBSCRIPT SMALL LETTER A
1D7B	e	LATIN SUBSCRIPT SMALL LETTER E
1D7C	o	LATIN SUBSCRIPT SMALL LETTER O
1D7D	x	LATIN SUBSCRIPT SMALL LETTER X
208F	/	SUBSCRIPT SOLIDUS

Subscripted <sub>a</sub>, <sub>e</sub>, and <sub>o</sub> are used to indicate the vowel coloring of a laryngeal **H**. This use is fairly common, perhaps second only to the use of **H** with subscript numbers, namely, **H**<sub>1</sub>, **H**<sub>2</sub>, and so on.

2. I see the need for four 'laryngeals':<sup>2</sup>

\*<sub>a</sub><sub>1</sub> = *H<sub>e</sub>*, perhaps [ʔ] when non-syllabic (> Hitt. *a-*, Gk. *ε* (etc.),<sup>3</sup>  
~ *ə*);

\*<sub>a</sub><sub>2</sub> = *H<sub>a</sub>*, probably [x] ~ [ħ] or [ʃ]<sup>4</sup> (> Hitt. *ḫ*(*ḫ*), Gk. *α*, Arm. *l*

\*<sub>a</sub><sub>3</sub> = *H<sub>o</sub>*, probably [x<sup>w</sup>] ~ [ħ<sup>w</sup>] or [ʃ<sup>w</sup>] (> Hitt. *ḫ*, Gk. *ο*, Arm. *?*)

\*<sub>a</sub><sub>4</sub> = *H<sub>a</sub>*, probably [ǰ] 'h'<sup>5</sup> (> Alb. *h-*, Gk. *α*).

**Figure 1.** Eric Hamp. 1989. "The Indo-European obstruent features and phonotactic constraints", in *The New Sound of Indo-European*, ed. Theo Vennemann. (Trends in Linguistics. Studies and Monographs; 41) Berlin & New York: Mouton de Gruyter.

Subscripted <sub>e</sub> is also used as a reduced vowel, as in **-<sub>e</sub>H<sub>2</sub>-** here:

To account for such Hittite forms  
by assuming a particular reduced grade (\*<sub>e</sub>H<sub>2</sub>-) or a 'vocalization' of  
\*H<sub>2</sub> > *a* does not seem to be possible, cf. §§ 41, 69, *Remark* below.

**Figure 2.** Fredrik Otto Lindeman. 1987. *Introduction to the 'Laryngeal Theory'*. Oslo: Norwegian University Press; distributed by Oxford University Press. Page 40.

Subscripted  $x$  is used after a laryngeal  $h$  to denote uncertain vowel colouring, as in  $*leu(h_x)$  and  $*h_xC(V)$ - here:

Another solid example of  $u < *eu$  is provided by Hitt. *luttāi*- 'window', which forms a word equation with Toch. B *lyauto* 'opening'. The root etymology of *lyauto* is given by Van Windekens, *le tokh.* I.266: PIE  $*leu(h_x)$ - 'cut, separate' (cf. Skt. *lunāti* 'cuts off'). Eichner, *MSS*

**Figure 3.** H. Craig Melchert. 1984. *Studies in Hittite Historical Phonology*. Göttingen: Vandenhoeck and Ruprecht. Page 59.

**Environment 1:**  $*h_xC(V)$ - ~  $*sh_xC(V)$ -,  
with (some) visible reflexes in the language groups already mentioned, enabling us to differentiate from cases of plain  $*(s)C$ -, with which  $*(s)h_xC$ - falls together in all the other daughters after the loss of the laryngeals. This assumes,

**Figure 4.** Mark Southern. 1999. *Subgrammatical Survival: Indo-European s-mobile and its Regeneration in Germanic*. (Journal of Indo-European Studies Monograph; 34) Washington, DC: Institute for the Study of Man. Page 93.

Subscripted  $l$  is used between two numbers or letters, as in  $*h_{1/3}al-u-t$ - and  $*dhug'hH_{(e/o)}ter$  here:

the wrong laryngeal:  $*h_{1/3}al-u-t]$

**Figure 5.** Mark Southern. 1999. *Subgrammatical Survival: Indo-European s-mobile and its Regeneration in Germanic*. (Journal of Indo-European Studies Monograph; 34) Washington, DC: Institute for the Study of Man. Page 124.

IE  $*-ǵ-$  (a combination of a  $*H$  with a reduced vowel) > SA  $*-a-$  in the middle of a word: HLuw. *tī-wali-tarali*- (daughter), Lyc. *kbatr-a*<sup>17</sup> < IE  $*dhug'hH_{(e/o)}ter$ , Skt. *dubitar*-, Gk. *θυγάτηρ*, Toch. B *tkācer*, oblique

**Figure 6.** Vyacheslav Ivanov. 2001. "Southern Anatolian and Northern Anatolian". in *Greater Anatolia and the Indo-Hittite Language Family: Papers presented at a colloquium hosted by the University of Richmond, March 18-19, 2000*, ed. Robert Drews, (Journal of Indo-European Studies Monograph; 38) Washington, DC: Institute for the Study of Man. Page 134.

## A. Administrative

### 1. Title

Proposal to encode five Indo-Europeanist phonetic characters in the UCS.

### 2. Requester's name

Deborah Anderson and Michael Everson.

### 3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

### 4. Submission date

2004-02-02

### 5. Requester's reference (if applicable)

### 6. Choose one of the following:

#### 6a. This is a complete proposal

Yes.

#### 6b. More information will be provided later

No.

## B. Technical -- General

### 1. Choose one of the following:

#### 1a. This proposal is for a new script (set of characters)

No.

**Proposed name of script**

**1b. The proposal is for addition of character(s) to an existing block**

Yes.

**1b. Name of the existing block**

Phonetic extensions, and Superscripts and Subscripts

**2. Number of characters in proposal**

5

**3. Proposed category (see section II, Character Categories)**

Category A.

**4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)**

Level 1.

**4b. Is a rationale provided for the choice?**

No.

**4c. If YES, reference**

**5a. Is a repertoire including character names provided?**

Yes.

**5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?**

Yes.

**5c. Are the character shapes attached in a legible form suitable for review?**

Yes.

**6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?**

Michael Everson.

**6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:**

Michael Everson, Fontographer.

**7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?**

Yes.

**7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?**

Yes.

**8. 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.

**9. 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/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.**

1D7A;LATIN SUBSCRIPT SMALL LETTER A;Ll;0;L;<sub> 0061;;;;N;;;;;

1D7B;LATIN SUBSCRIPT SMALL LETTER E;Ll;0;L;<sub> 0065;;;;N;;;;;

1D7B;LATIN SUBSCRIPT SMALL LETTER O;Ll;0;L;<sub> 006F;;;;N;;;;;

1D7B;LATIN SUBSCRIPT SMALL LETTER X;Ll;0;L;<sub> 0078;;;;N;;;;;

208F;SUBSCRIPT SOLIDUS;Po;0;ES;<sub> 002F;;;;N;;;;;

## **C. Technical -- Justification**

**1. Has this proposal for addition of character(s) been submitted before? If YES, explain.**

No.

**2a. 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.

**2b. If YES, with whom?**

H. Craig Melchert, U of North Carolina, Chapel Hill; Brent Vine, UCLA; Andrew Garrett, UC Berkeley.

**2c. 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?**

Linguists and phoneticians.

**4a. The context of use for the proposed characters (type of use; common or rare)**

Indo-European phonetic reconstructions.

**4b. Reference**

As above.

**5a. Are the proposed characters in current use by the user community?**

Yes.

**5b. If YES, where?**

See publications above.

**6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?**

Yes.

**6b. If YES, is a rationale provided?**

Yes.

**6c. If YES, reference**

Keep with other subscripts used for phonetic purposes

**7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?**

**8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?**

No.

**8b. If YES, is a rationale for its inclusion provided?**

**8c. If YES, reference**

**9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?**

No.

**9b. If YES, is a rationale for its inclusion provided?**

**9c. If YES, reference**

**10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?**

No.

**10b. If YES, is a rationale for its inclusion provided?**

**10c. If YES, reference**

**11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?**

No.

**11b. If YES, is a rationale for such use provided?**

**11c. If YES, reference**

**12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?**

No.

**12b. If YES, reference**

**13a. Does the proposal contain characters with any special properties such as control function or similar semantics?**

No.

**13b. If YES, describe in detail (include attachment if necessary)**

**14a. Does the proposal contain any Ideographic compatibility character(s)?**

No.

**14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?**