

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 six 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****Replaces: N2705****Date: 2004-06-07**

A number of subscripted letters are used in Indo-European linguistic materials, though their use is not necessarily limited to Indo-European. While a few Latin letters are currently encoded as modifier letters, the needs for Indo-European extend beyond this set. Six characters are proposed here. Styled text is not seen as appropriate for these; Indo-Europeanists already make use of the subscript digits _{1 2 3 4}, and superscript ^h and ^w 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
1D7E	_ə	LATIN SUBSCRIPT SMALL LETTER SCHWA
208F	/	SUBSCRIPT SOLIDUS

2. I see the need for four ‘laryngeals’:²

*_{a1} = *H_e*, perhaps [ʔ] when non-syllabic (> Hitt. *a-*, Gk. *ε* (etc.)),³
~ *ə*);

*_{a2} = *H_a*, probably [x] ~ [h] or [ʃ]⁴ (> Hitt. *ḫ*(*ḫ*), Gk. *α*, Arm. *ի*

*_{a3} = *H_o*, probably [x^w] ~ [h^w] or [ʃ^w] (> Hitt. *ḫ*, Gk. *ο*, Arm. *ո*),

*_{a4} = *H_a*, probably [a] ‘h’⁵ (> Alb. *h-*, Gk. *α*).

Figure 1. Subscripted _a, _e, and _o are used to indicate the vowel coloring of a laryngeal **H** (or **h** as in figures 8 and 9 below). This may be used instead of the more common **H** with subscript numbers, namely, **H₁**, **H₂**, and so on; subscript digits are shown with schwa in this example.

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.

	<i>h</i> ₁	<i>h</i> ₂	<i>h</i> ₃	
Laryngeals	[^ʔ ; ^g ə̯, ə̯ ^g] e e	[h; h ^g ə̯, ə̯ ^g h] a a	[^ç ; ^g ə̯, ə̯ ^ç] o o	ʔisə̯ (ʔ/h) e/a

Figure 2. Subscripted *e*, *a*, and *o* are used along with subscript *l* in plain (non-italic) phonetic text transcription from the TITUS-Projekt.

<http://titus.fkidg1.uni-frankfurt.de/didact/idg/idgphon.htm#XEN9>, “Phonemes and their phonetic realization (overview)”

To account for such Hittite forms
by assuming a particular reduced grade (*-eH₂-) or a ‘vocalization’ of
*H₂ > a does not seem to be possible, cf. §§ 41, 69, Remark below.

Figure 3. Subscripted *e* is also used as a reduced vowel, as in -eH₂- here.

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

confirm this statement. On page 288, for example, he insists that there were three ‘reduced’ vowels *e a o*; from this conclusion he could readily move to any of the theories on laryngeals. But he refuses to admit any connection between *e a o* and the long vowels of root syllables. Hence his analysis is seriously deficient.

Figure 4. Subscripted *e*, *a*, and *o* are used as reduced vowels here.

W. P. Lehman. 1958. “On earlier stages of the indo-European nominal inflection”, in *Language* 34.2:181.

It developed in patterns when the subject directly preceded a finite verb. Conditions were then right for lengthened grade; for with the loss of accent on a following verb, its stem vowel dropped out and the preceding vowel was lengthened; for example, pre-Indo-European /p_etér/ + /éyty/ became /p_etér yty/.

Figure 5. Subscripted *e* is also used as a reduced vowel, as in p_etér here.

W. P. Lehman. 1958. “On earlier stages of the indo-European nominal inflection”, in *Language* 34.2:198.

development of early Indo-European. In the weakening of the original sequences *eR oR*, where *R* stands for *r l m n*, the first stage was *eR oR*, which later coalesced in *oR*. For this reason in the perfect tense, for example, s.*wert- was opposed to pl. *w_ort- (not *w_ert-). The over-short *o* was, however, perceived as belonging to the phoneme *o*, with the result that *w_ort- arose to replace the singular *wert-. By the same

Figure 6. Subscripted *o* and *e* are used as reduced vowels, as in *w_ort- and *w_ert- here.

O. Szemerényi, *Introduction to Indo-European Linguistics*, p. 120 (Oxford, 1990).

Even before Myc. *i-po-no* '(cooking) bowls, dutch ovens' roiled these waters,³³ all was not clear with *ἰπυός* in formal terms (even if the semantic development of the term within Greek can be satisfactorily traced, on which see now Chadwick 1996: 161ff.): among other problems, it was necessary to operate with one or another irregular treatment of a zero-grade **uk^w-nó-* (itself not ideal for a root with *a*-vocalism), such as a dissimilation of **u-* to **i-* conditioned by the following labiovelar, or else a Güntertian schwa-secundum formulation in terms of a *"*w_jk^w-nó-*" (Güntert 1916: 25, among forms allegedly displaying a development *"a [ə = "schwa secundum" — add. BV] zu gr. ι bei benachbartem u, u"*, cf. n. 21 above). The vowel-dissimilation account founders on the likelihood that in

Figure 7. Subscripted _o is used to denote a reduced vowel, as in **w_ok^w-nó-* here.

Brent Vine. 1999. "Greek rhiza 'root' and 'Schwa Secundum'," in *UCLA Indo-European Studies* 1, ed. by Brent Vine and Vyacheslav Ivanov.

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. *lanāti* 'cuts off'). Eichner, *MSS*

Figure 8. Subscripted _x is used after a laryngeal **h** to denote uncertain vowel colouring, as in **leu(h_x)* here.

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 9. Subscripted _x is used after a laryngeal **h** to denote uncertain vowel colouring, as in **h_xC(V)-* here.

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.

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

Figure 10. Subscripted _j is used between two numbers indicating the laryngeal, as in **h_{1/3}al-u-t-* here.

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 **-ə-* (a combination of a **H* with a reduced vowel) > SA **-a-* in the middle of a word: HLuw. *nī-wali-tarali-* (daughter), Lyc. *kbasr-a¹⁷* < IE **dhug'hH(e_o)ter*, Skt. *dubitar-*, Gk. *θυγάτηρ*, Toch. B *tkācer*, oblique

Figure 11. Subscripted _j is used between two letters indicating the laryngeal, as in **dhug'hH(e_o)ter* here.

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.

udu-aslum_x: the transliteration aslum_x for a-lum is based on ancient lexical speculation of doubtful value (but cf. Steinkeller, "Sheep and goat terminology in Ur III Sources from Drehem," *BSA* 8 [1995]: 52). The Akkadian word *aslu* was used in late standard Babylonian, especially Assyrian royal inscriptions of the Neo-Assyrian period. It is also unclear why an Akkadian word *aslum* would appear in the syllabic spelling a-lum. The word a-lum seems Akkadian. Udu-a-lum may mean "city sheep" or, if derived from *ālu* II, "ram sheep." But we do not know the reading of the sign LUM in this word and the reading a-lum is simply a guess.

Figure 12. Subscripted _x is used in an Cuneiformist context as in **udu-aslum_x** here.

A. Administrative

1. Title

Proposal to encode six 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

2004-06-07

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

6

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;;;;;

1D7C;LATIN SUBSCRIPT SMALL LETTER O;Ll;0;L;<sub> 006F;;;N;;;;;

1D7D;LATIN SUBSCRIPT SMALL LETTER X;Ll;0;L;<sub> 0078;;;N;;;;;

1D7E;LATIN SUBSCRIPT SMALL LETTER SCHWA;Ll;0;L;<sub> 0259;;;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?