# ISO/IEC JTC1/SC2/WG2 N2514

2002-11-10

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 one Latin letter in the UCS

**Source:** Michael Everson

**Status: Individual Contribution** 

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2002-11-10

N2457, a contribution from the US National Body, proposed two double diacritics used widely in US lexicographical practice, in order to ensure that such data could be encoded in the UCS. One character from that tradition remains to be encoded.

# 1D6B **U** LATIN SMALL LETTER UE

3 day, fade, date, gorta	00 (in transcriptions of foreign words only) indicates
a bother, car; most American speakers have the same vowel in father, cart	a vowel with the quality of a in bone but longer, not the sound of so in food: comte \koonto.
<ol> <li>father as pronounced by speakers who do not rhyme it with bother; farther and cart as pro- nounced by r-droppers</li> </ol>	Dpepper, lip
	I rarity, one pronunciation of tar
22 bad, bag, fan as often pronounced in an area	\$source, less
having New York City and Washington, D. C., on its perimeter; in an emphatic syllable, as be- fore a pause, often \aaa\	Sh with nothing between, as in Shy, mission, ma- chine, special (actually, this is a single sound, not
21, as in some pronunciations of bag, bang, pass	two); with a stress mark between, two sounds as in death's-head \'deths,hed\
a.unow, loud, some pronunciations of talcum	t tic, attack; one pronunciation of latter (alterna-
bbaby, rib	tive \d-\}
ChChin, nature \'nāchə(r)\ (actually, this sound is \\\\ + \sh\)	th with nothing between, as in thin, ether (actualty, this is a single sound, not two); with a stress mark between, two sounds as in knighthood \n\cdot\n\cdo\n\cdot\n\cdot\n\cdot\n\cdot\n\cdot\n\cdot\n\cdot\n\cdot\n\cdot\n
delder, undone	th then, either (actually, this is a single sound, not
d as in the usual American pronunciation of latter.	two)
ladder	Ürule, fool, youth, union \'yllay≥a few \'fytt\
Č bet, bed	. pull, wood, curable \'kyūrabal\
c, c beat, nosebleed, evenly, sleepy	UC German füllen, hübsch
as in one pronunciation of evenly, sleepy, envious, ignoous (alternative \i\)	Erench rite, German stillen
(in transcriptions of foreign words only) indicates a wowel with the quality of e in bet but tong, not	W We, away
the sound of ee in sleep: en arrière \annaryeer\	Y yard, cue \'ky@ union \'yunyan\
eu as in one pronunciation of elk, helm	. J, Juca, our t hju ti amon t janjant

# A. Administrative

#### 1. Title

Proposal to encode one Latin letter in the UCS.

2. Requester's name

Michael Everson

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

Individual contribution.

4. Submission date

2002-11-10

- 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

Ves

1b. Name of the existing block

2. Number of characters in proposal

1

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?

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. Gove, Philip Babcock, ed. 1993 (1961). Webster's Third New International Dicitonary of the English Language, Unabridged. Cologne: Könemann. ISBN 3-8290-5292-8

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.

Spacing letter, category "Lo", bidi category "L" (strong left to right)

1D6B

## 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.)?

No.

2b. If YES, with whom?

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?

Dictionary users.

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

Phonetic transcription.

4b. Reference

Merriam-Webster dictionaries.

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

Yes.

5b. If YES, where?

By dictionary publishers.

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

Latin phonetic character.

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. While conventional font ligation can produce a glyph similar to <u>, this letter is a unitary character with a precise phonetic value, as <u> and <u> are in the IPA tradition.

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?

Nο.

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)?

Nο

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