

**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 add the OBLIQUE HYPHEN to the UCS
Source: Michael Everson
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2021-01-12

1. Introduction. The UCS contains a number of hyphen characters used for various purposes. This proposal requests one OBLIQUE HYPHEN which is used generally in medieval texts and in particular in a pairing with the existing DOUBLE HYPHEN in order to indicate the end of a hyphenated word. If this proposal is accepted, the last character in the list below will be added alongside the other three.

- 002D HYPHEN-MINUS
- ≧ 2E17 DOUBLE OBLIQUE HYPHEN
- = 2E40 DOUBLE HYPHEN
- ↯ 2E5D OBLIQUE HYPHEN *proposed here*

2. Use in the Middle Ages and Renaissance. In medieval European manuscripts, a line-breaking hyphen was typically oblique, either single *˘* or double *≧*. In Murdoch Nisbet’s translation of the New Testament into Middle Scots, the two are used alongside one another, the double oblique hyphen at the line-break itself, and the single oblique hyphen at the end of the broken word as a delimiter of the hyphenation. In other manuscripts of the medieval period the end-of-line oblique hyphen can be either single or double; in the Elizabethan period a single OBLIQUE HYPHEN had become the norm. See Figures 1–3.

3. Modern use. The Oxford University Press distinguishes horizontal HYPHEN-MINUS from OBLIQUE HYPHEN explicitly to distinguish end-of-line soft hyphens and end-of-line hard hyphens which should be retained as they are permanent orthographic elements in a word. In Figure 4 below, the Oxford use of OBLIQUE HYPHEN are shown alongside some other conventions other publishers have made use of from time to time. To these a few others can be added from the Wikipedia article on DOUBLE HYPHEN.

Modern Standard	Oxford	Tilde	Two hyphens	Merriam-Webster	Nisbet
cross-	cross-	cross~	cross-	cross≧	cross≧
country	country	country	-country	country	country-

4. Linebreaking. Line-breaking properties for this are suggested as follows.

Like U+2E40 = DOUBLE HYPHEN (BA (Break_After)): 2E5D

5. Unicode Character Properties. Character properties are proposed here.

2E5D;OBLIQUE HYPHEN;Pd;0;ON;;;;;N;;;;;

6. Bibliography.

- Roberts, Rosemary, ed., et al. 2005. *New Hart's Rules: Adapted from The Oxford Guide to Style by R. M. Ritter*. Oxford: University Press. ISBN 978-0-19-861041-0
- Tannenbaum, Samuel A. 1931. *The handwriting of the Renaissance: Being the Development and Characteristics of the Script of Shaksperes Time*. London: George Routledge & Sons; New York: Columbia University Press.
- Wright, C. E. 1960. *English vernacular hands from the twelfth to the fifteenth centuries*. (Oxford Palaeographical Handbooks.) Oxford: Clarendon Press.

7. Figures

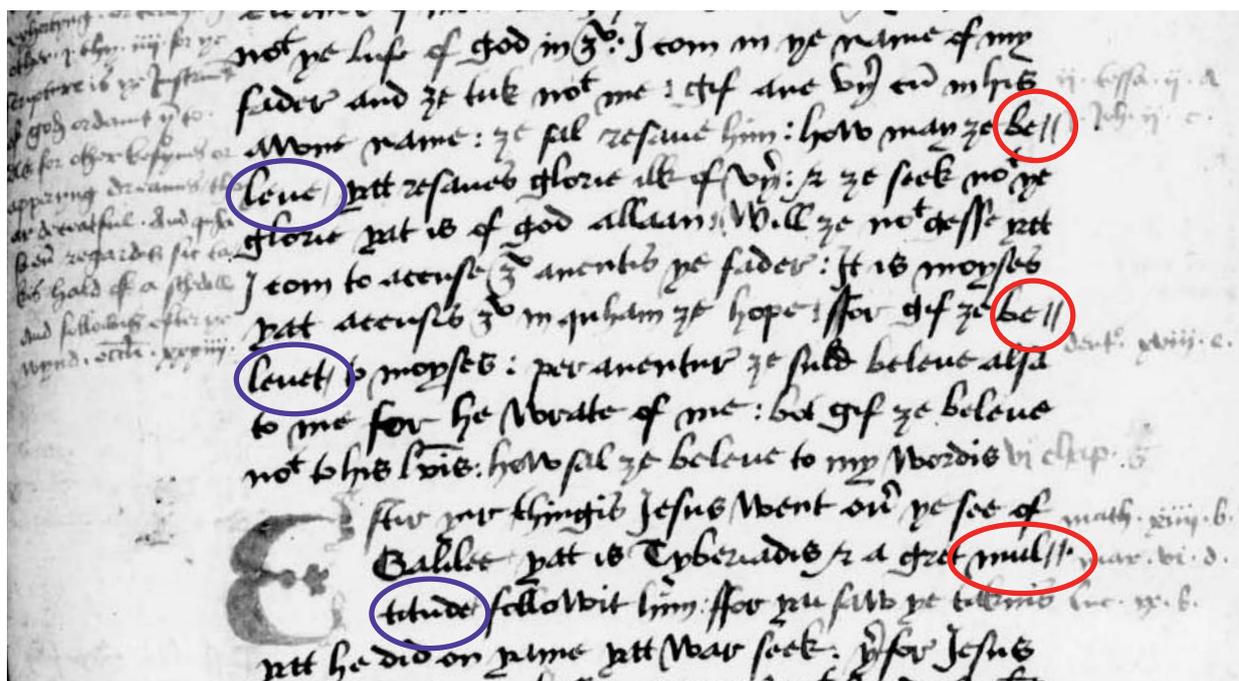


Figure 1. In the Gospel of John 5:43–6:2 (Egerton MS 2880, f. 91v). In four three instances words are hyphenated with DOUBLE OBLIQUE HYPHEN and the hyphenation sequences ended by the OBLIQUE HYPHEN. The palaeographic transcription of this passage (with punctuation and quotation marks added) follows:

no^t þe lufe of god in ȝ: ⁴³I com in þe name of my
 fader^ȝ and ȝe tuk no^t me: gif ane v^þ cū in his
 awne name: ȝe fal ȝefaue him: ⁴⁴How may ȝe be=
 leue- þat ȝefaues glozie ilk of v^þ: 7 ȝe feek no^t þe
 glozie þat is of god allaan: ⁴⁵Will ȝe no^t geffe, þat
 I com to accuse ȝ anentis þe fader^ȝ: It is Moyfes
 þat accusis ȝ in quham ȝe hope: ⁴⁶For gif ȝe be=
 leuet- to Moyfes: perauentu^ȝ ȝe fuld beleue alla
 to me for he wrate of me: ⁴⁷bot gif ȝe beleue
 no^t to his l^ris: how fal ȝe beleue to my wordis *vi chap.*
⁶⁻¹Eftir þir thingis Iefus went ou^ȝ þe see of
 Galilee þat is Tybeziadis ²7 a gzet mul=
 titude- followit him: For þai saw þe taknis
 þat he did on þame þat war feek: ³þfor Iefus ...



O my wel beloved in our lord god my nicholas
 mayster of ye order of sempyngha Peysby-
 wherch ordre is entytled on to ye
 name of seynt gilbert Jffrey J. C. John
 a moung^r doctouris left sende ye capgrave.
 uerens as to stherch dignyte. desiring clenesse
 to your soule and helth to your body noll
 witht myc felbe dyes was notyfyed on
 to me yat ye luf of our fader seynt augu-
 styn wherch yat J truisalat in to our tinge
 at instans of a cteym woma was brobt
 to your presens wherch lyke 30ll wel as it
 is told saue ze wold J schul adde yto alle
 yo religyous yat lyue vnder his reule
 but to ye Jansber yat it was not my char-
 ge but if men like for to knowe ys mat^r
 diffnsely yei may lerne it in a bmo yat seid
 at cambryg^r ye zer be for myn opposicion
 wherch fmon vmp^rhap J whil sette in englysch
 in ye list ende of ys werke ehan astur ze
 thid red ys luf of seynt augustty ze sayde
 to on of my frendes yat ze desired gretly
 ye luf of seynt gilbert schuld be traslat
 in the same forme Thus mad he instance
 to me and J granted both your peticion for
 J wold not frustate hym of his mediacou
 to ye honom of god and of all seyntis ya
 whil we kepyne ys trefys nymelich for the
 solitaryo. Some of your religion wherch in

Figure 2a. Text of the Middle English translation of *The Lives of St Augustine of Hippo and St Gilbert of Sempringham*, British Museum Additional MS 36704, f. 46r, from Wright 1960:21. Instances of the OBLIQUE HYPHEN have been circled. The transcription in Figure 2b shows a hyphen in “augustyn” though there does not appear to be one on the page. They are all faint (as is usual).

21. London, British Museum, Additional MS. 36704

LIVES OF ST. AUGUSTINE OF HIPPO AND ST. GILBERT OF SEMPRINGHAM, &c., by John Capgrave, Austin Friar of King's Lynn, Norfolk (d. 1464), author also of the 'Liber de Illustribus Henricis' and a Chronicle of England (to 1417).

Written A.D. 1451 (the year of the composition of the Life of St. Gilbert, cf. fol. 116^r) probably by Capgrave himself (the hand agreeing with that in Cambridge University Library MS. Gg. 4. 12). (On the question of Capgrave's *autograph* MSS. see H. M. Bannister in *Ye Solace of Pilgrimes: A Description of Rome, circa A.D. 1450 by John Capgrave*, ed. C. A. Mills (Oxford, 1911), pp. xv-xxii.)

Paper: 20.5 × 14.3 cm.

Handwriting is neat and regular but individual and not calligraphic; corrections are in same hand as text.

Note. *th* is used in place of *þ* for capitals and at the end (and sometimes the beginning) of words; *þ* itself appears in the degenerate 'wyn' form; the 2-form of *r* is not used; *ȝ* is used for only *y*-sound (e.g. 'ȝour'); *g* itself is written like *y* with a cross-bar added (*ȝ*); long *s* except finally, when round *s* is used (the latter characteristically in Capgrave's hand rather like a small-size *B*; cf. also pl. 16); *y* is sometimes dotted.

Abbreviations: it is a question whether the ' after *r* is for 're' or whether it has become now a meaningless scribal flourish, nor is it certain whether the '-es' symbol (*es*) after *g* in 'among' (l. 5) is to be extended or not. Again, in words such as 'petyciou' (e.g.) it is doubtful whether the scribe had in mind *ou* or *on* when he wrote the last two letters.

Text printed in: *John Capgrave's Lives of St. Augustine and St. Gilbert of Sempringham*, ed. J. J. Munro (E.E.T.S., o.s., no. 140, 1910).

Also reproduced with another page in: New Pal. Soc., ser. 1, pl. 70.

Our plate shows fol. 46^r.

TRANSCRIPTION

TO my wel beloued in our' lord god
 maystir of þe order' of s̄mp̄yghām
 wech ordre is entytled on to þe
 name of seynt gilbert j ffrer' I. C.
 a monge doctouris lest sende re'
 uerens as to swech dignyte desirig clēnesse
 to ȝour' soule and helth to ȝour' body Now
 with inne fewe dayes was notyfyed on
 to me þat þe lyf of our' fader seynt augus'
 tyn wech þat j translat in to our' tūge
 at instaūs of a c'teyn womā was browt
 to ȝour' presens wech lykyd ȝow wel as it
 is told saue ȝe wold j schul adde þ'to alle
 þoo relygyous þat lyue vndyr his reule
 But to þis j answer' þat it was not my char'
 ge but if men like for to knowe þis mater'
 diffusely þei may lerne it in a s'mō þat seid
 at cambrig' þe ȝer' befor' myn opposiciōū
 wech s'mon vnphap j wil sette in englich
 in þe last ende of þis werk Than aftir ȝe
 had red þis lyf of seynt augusty ȝe sayde
 to on of my frendes þat ȝe desired gretly
 þe lyf of seynt gilbert schuld be translat
 in the same forme Thus mad he instaūce
 to me and j graūted both ȝour' petyciōū for j ē his
 j wold not frustrate him of his mediaciōū
 To þe honour' of god and of al seyntis þā
 wil we begyne þis tretys namelych for the
 solitarie womē of ȝour' religiōū wech vn'

M. Nicholas
 Reysby

Figure 2b. Transcription of the Middle English translation of *The Lives of St Augustine of Hippo and St Gilbert of Sempringham*, British Museum Additional MS 36704, f. 46r, from Wright 1960:21. Instances of the OBLIQUE HYPHEN can be seen, as well as a number of instances of the COMBINING OVERCURL, here transcribed as an apostrophe.

A *hyphen* at the end of a line was not often employed. Penmen and printers divided syllables in any way they pleased, as in *agains-t*, without a hyphen. When the hyphen was used, it usually consisted of two short, parallel, horizontal or oblique strokes with or without a dot; in the thirteenth and fourteenth centuries it consisted of a single short oblique stroke. Compound words, as “craftsman,” “free-born,” were generally written as two words, without a hyphen.

Figure 3. Discussion of the history of hyphenation from Tannenbaum 1931:146. The text is to introduce readers to the realities of Elizabethan manuscripts (which can be pretty tricky to read). Tannenbaum points out that two strokes were often used in the earlier period, but in the 14th and 15th centuries a single OBLIQUE HYPHEN came to be the norm.

end of a line (in unjustified setting). The narrower the column, the more necessary this becomes. Some divisions are better than others, and some are unacceptable because they may mislead or confuse the reader. Rules governing division are based on a combination of a word's construction (i.e. the parts from which it is formed) and its pronunciation, since exclusive reliance on either can yield unfortunate results. The following offers general guidance only; for individual cases, consult the *New Oxford Spelling Dictionary*. See also 2.5.2 for a discussion of the general principles of page layout and proof-reading. For word division in foreign languages, see Chapter 12, under the languages concerned.

A hyphen is added where a word is divided at the end of a line. This is known as a **soft hyphen** or **discretionary hyphen**:

con-
trary

If a word with a hard hyphen is divided after its permanent (keyed) hyphen, no further hyphen is added:

well-
developed

In most texts the hyphens in the examples above (*con-trary* and *well-developed*) will use the same symbol (-). Sometimes, as in dictionaries or other reference works in which it is important for the reader to know whether an end-of-line hyphen is a permanent one or not, a different symbol, such as *is -* (as in this book), is used when words are divided:

con-
trary

A tilde (~) is also occasionally used:

con~
trary

Formerly, a permanent hyphen was sometimes repeated at the start of the following line, thus:

well-
-developed

but this is rare nowadays.

In copy to be keyed, add a stet mark (from Latin, 'let it stand', meaning that the original form should be retained) to any permanent (hard) hyphen that falls at the end of a line, to indicate that it must be keyed.

Figure 4. Sample from Roberts et al. 2005:58 showing the modern use of a single OBLIQUE HYPHEN (circled in red) to indicate a hyphen which is a permanent orthographic element in a word. The practice of some other dictionaries is also given. Note that “con-trary” is given by the Oxford editor as an example, though that word in an end-of-line context would not have a hard hyphen.

A. Administrative

1. Title

Proposal to add two mediaeval punctuation characters to the UCS

2. Requester's name

Michael Everson

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

Individual contribution.

4. Submission date

2021-01-12

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.

1b. Proposed name of script

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

Yes

1d. Name of the existing block

Supplementary Punctuation

2. Number of characters in proposal

1.

3. Proposed category (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)

Category B.1.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. 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)?

Yes.

8. 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.

See above.

C. Technical – Justification

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

Yes, some of the characters have. See N3193.

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?

The authors are members of the user community.

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?

Medievalists, Latinists, and other scholars.

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

Used historically and in modern editions.

4b. Reference

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

Yes.

5b. If YES, where?

Scholarly publications.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Accordance with the Roadmap. Keep with other punctuation characters.

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

No.

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?

Yes.

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

Yes.

10c. If YES, reference

Some characters are ancestors of modern characters.

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

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?