

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 two characters for Medieval Cornish to the UCS



Source: Michael Everson

Status: Liaison Contribution

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

Date: 2017-09-27

0. Summary. This proposal requests the encoding of one combining mark and one punctuation characters. If this proposal is accepted, the following characters will exist:

	1DFA	COMBINING FINAL SWASH	<ul style="list-style-type: none"> • used in medieval Cornish • fuses typographically with a, e, m, n, r, u, y
	2E4F	DOUBLE OBLIQUE HYPHEN WITH FALLING DOTS	<ul style="list-style-type: none"> • used in medieval Cornish

1. COMBINING FINAL SWASH. Medieval handwriting in Cornish uses a variety of mechanisms for representing an abbreviation of *m* and *n*. These mechanisms occur alongside one another and can all be represented by characters in the UCS (U+0306 COMBINING OVERLINE, U+0311 COMBINING INVERTED BREVE, and U+0352 COMBINING FERMATA).

\bar{m} \hat{m} \hat{m} \bar{n} \hat{n} \hat{n}

These are read *mm* and *nn*; *mn* and *nm* are in principle possible but would be extremely rare. These marks also appear on vowels:

\bar{a} \hat{a} \hat{a} \bar{y} \hat{y} \hat{y}

Which can be read *am* and *ym* or *an* and *yn*. But one kind of abbreviation, which is relatively productive, is also found:

\bar{a} \bar{e} \bar{m} \bar{n} \bar{r}/\bar{r} \bar{u} \bar{y}

These forms are polyvalent. This mark may simply be a meaningless swash form, or it may be an abbreviation. Thus the readings for these may be *a, e, m, n, r, u, y*, or *am, em, mm, nm, rm, um, ym*, or *an, en, mn, nn, rn, un, yn*. There is no way of telling without knowing the reading rules for these. But a palaeographic representation of the text is impossible for these last forms without the combining character proposed here. We already have a similar analogous situation with U+035B COMBINING ZIGZAG ABOVE, which was encoded as an abbreviation representing *er* and *re*. Here are some examples:

war^ʼ dr^ʼ der^ʼ ma^ʼne

These are forms attested in the manuscripts: the first two words are Middle Cornish *war* ‘on’ (where the COMBINING ZIGZAG is otiose) and *dre* ‘through’ (where it represents *-e*), and the second two are Middle English *dere* ‘dear’ (where the ZIGZAG represents *-e*), *manere* ‘manner’ (where it represents *-er-*). The encoded COMBINING ZIGZAG can be used to correctly represent the text, whether the reading of it is *e* or *re* or *er*, or whether it is just an otiose swash. Similarly, when we have a word ending in *-ð*, we do not know whether it is *-en* or *-em* or *-e*—but with the new COMBINING FINAL SWASH it is possible to represent the text accurately regardless of the meaning.

It is usual in the UCS that diacritics that fuse typographically with base characters are encoded atomically, but since this is a productive abbreviation character not used in a standard orthography it is considered reasonable to encode it as a single combining character. An informative note listing the characters which have been observed making use of it is recommended for the names list.

2. DOUBLE OBLIQUE HYPHEN WITH FALLING DOTS. This separator character has only been observed in Cornish manuscripts, and there is some possibility that it was devised at the scriptorium at Glasney, where Cornish orthography is thought to have been developed and maintained. The character is used with some regularity between verses as a simple separator, and cannot be represented by any other punctuation characters in the standard. It has been found so far in the 14th-century poem *Pascon agan Arluth* ‘The Passion of our Lord’, in the drama *Beunans Meriasek* ‘The Life of St Meriasek’ (1504), and in the recently-discovered drama *Bewnans Ke* ‘The Life of St Ke’. In the best (and earliest) hands the glyph is definitely two oblique strokes flanked by two angled dots, though in weaker hands glyph variants can be observed. These could be represented by sequences of existing characters if necessary, but the exemplar character cannot be. It is certain that only one character is needed.

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

1DFA: CM (Combining Mark)

2E4F: BA (Break_After)

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

```
1DFA;COMBINING FINAL SWASH;Mc;210;L;;;;N;;;;;  
2E4F;DOUBLE OBLIQUE HYPHEN WITH FALLING DOTS;Po;0;ON;;;;N;;;;;
```

6. Bibliography.

The following forthcoming volumes in the series Corpus Textorum Cornicorum require the characters proposed in this document.

Williams, Nicholas, Alan M. Kent, & Michael Everson. Forthcoming. *The Charter Fragment and Pascon agan Arluth*. (Corpus Textorum Cornicorum; 1) Portlaoise: Everttype.

Williams, Nicholas, Alan M. Kent, & Michael Everson. Forthcoming. *Beunans Meriasek*. (Corpus Textorum Cornicorum; 5) Portlaoise: Everttype.

Williams, Nicholas, Alan M. Kent, & Michael Everson. Forthcoming. *Bewnans Ke*. (Corpus Textorum Cornicorum; 6) Portlaoise: Everttype.

Williams, Nicholas, Alan M. Kent, & Michael Everson. Forthcoming. *Gwryans an Bÿs (The Creacon of the World)*. (Corpus Textorum Cornicorum; 8) Portlaoise: Everttype.

Figures

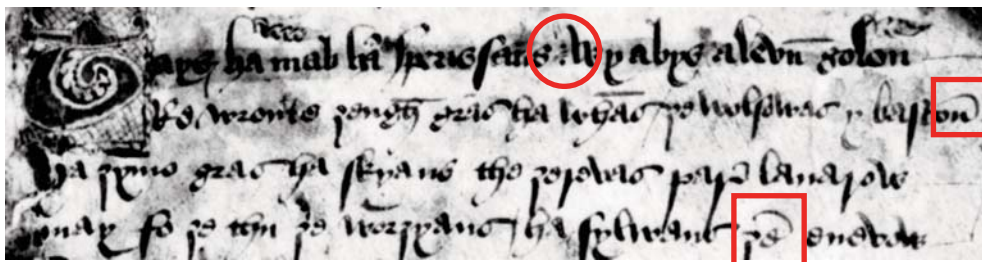


Figure 1. The first line of *Pascon agan Arluth* ‘The Passion of our Lord’ (BL MS Harley 1782B, fol. 1r), showing DOUBLE OBLIQUE HYPHEN WITH FALLING DOTS and COMBINING FINAL SWASH on *n* (where it is otiose) and *e* (where it is meaningful). The text reads, in palaeographic presentation, normalized text, and translation:

Tays ha mab hâ fpezis fans ð 6y abys a levñ golon
 Re wzonts zeugh gras ha whās / ze wolfowas y bascon
 Ha zymo gras ha skyans the zerevas par' lauarot
 may fo ze thu ze wozyans / ha fylwans 3@ enevo

*Tas ha Mab ha'n Spyrys Sans,
 why a bys a leun-golon,
 re wrauntyo dhywegh grâss ha whans
 dhe wolsowes y Bassyon;
 ha dhymmo grâss ha skians
 dhe dherivas par' dell wòn,
 may fo dhe Dhuw dh'y wordhyans
 ha selwans dhe'n Gristenyon.*

May Father, Son and the Holy Spirit—
 you who pray from the bottom of your heart—
 grant you grace and yearning
 to listen to his Passion,
 and to me grace and wisdom
 to recount as well as I can,
 that it may be for the glory of God
 and the salvation of Christians.

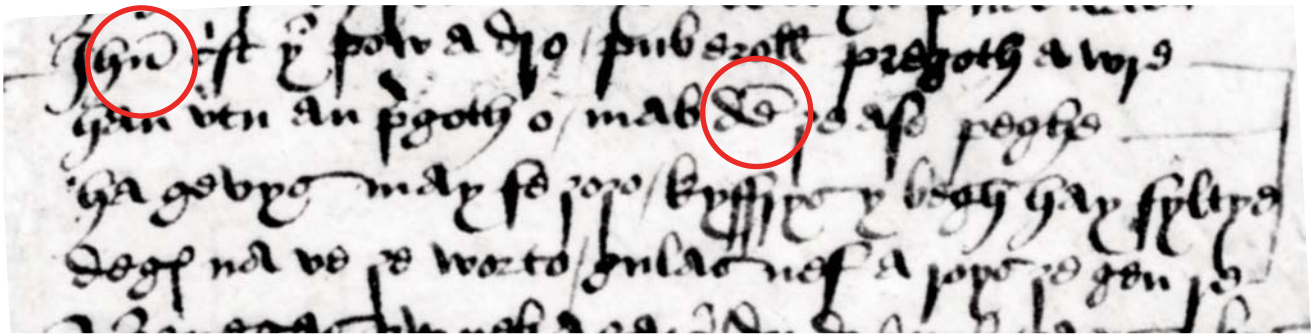


Figure 2. Verse 23 of *Pascon agan Arluth* ‘The Passion of our Lord’ (BL MS Harley 1782B, fol. 3r), showing COMBINING FINAL SWASH on *u* and *e*. On *u* the swash is decorative but it is meaningful on the *e* (where it means *n*). The text reads, in palaeographic presentation, normalized text, and translation:

Ihū cft ŷ pow a dro / pub e2oſſ pꝛegoth a wre
 han vtu an pꝛgoth o / mab dē 3e afe peghe
 ha gevys may fe 3o3o / kyffrys y begħ hay fyltye
 degſ na ve 3e wortto / gulas nef a roys 3e gen re

*Jesu Crist i'n pow adro
 püb eur oll pꝛegoth a wre;
 ha vertu an pꝛgoth o
 mab den dhe asa peha,
 ha gevys may fe dhodho
 keffrys y begħ ha'y fylta,
 degys na ve dhyworto
 gwlas nev ha rës dhe gen re.*

Jesus Christ all around the country
 used always to preach;
 and the essence of the preaching was
 that man should give up sinning,
 so that there should be forgiven him
 both his sin and his corruption,
 to the end that the kingdom of heaven should not
 be taken from him and given to others.

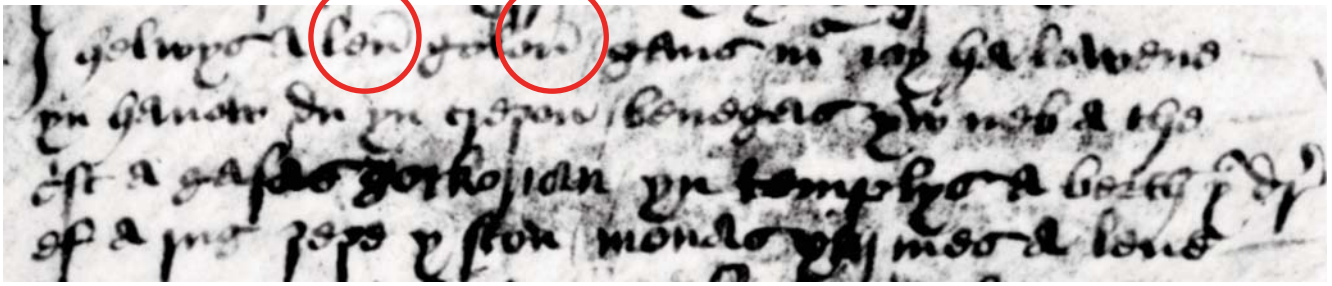


Figure 3. Verse 30 of *Pascon agan Arluth* ‘The Passion of our Lord’ (BL MS Harley 1782B, fol. 3v), showing COMBINING FINAL SWASH on *u* and *n*. On *n* the swash is decorative but it is meaningful on the *u* (where it means *n*). The text reads, in palaeographic presentation, normalized text, and translation:

I helwys a leū golon / gans m̄ ioŷ ha lowene
 yn hanow du yn trezon / benegas yw neb a the
 c̄ft a gafas gozkorian / yn templys a bezt̄h ŷ dr̄
 ef a rug zeze y fcon / monas yn mes a lene

*Y helwys a leun-golon
 gans meur joy ha lowena
 “In hanow Duw intredhon
 benegys yw neb a dheu!”
 Crist a gafas gwycoryon
 i’n templys aberth i’n dre.
 Ev a wrug dhedha yn scon
 mones in mes alena.*

People called out from the bottom of their heart
 with great joy and gladness,
 “In the name of God among us
 is blessed he who comes!”
 Christ found traders
 in the temples within the city.
 Quickly he caused them
 to depart away from there

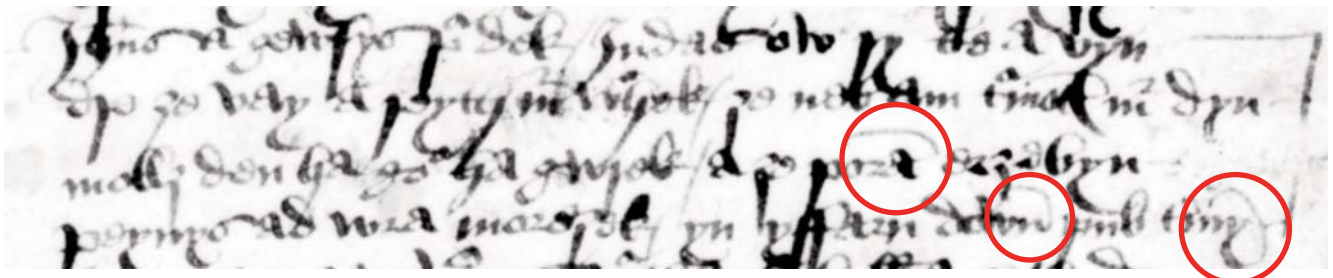


Figure 4. Verse 66 of *Pascon agan Arluth* ‘The Passion of our Lord’ (BL MS Harley 1782B, fol. 6v), showing COMBINING FINAL SWASH on *a*, *n* and *y*. On *n* the swash is decorative but it is meaningful on the other two (it means *n* on both). The text reads, in palaeographic presentation, normalized text, and translation:

Ihus a gewfys þ̄ dek / Iudas ow ry te a vyn
 dre ʒe vay a reyth m̄ whek / ʒe neb am t̄môt m̄ dyn
 mollʒ den ha gō ha gwrek / a ʒe poʒā eʒʒebyn
 peynys ad wʒa moʒeʒek / yn yffaʒn down̄ pub t̄m̄ȳ

*Jesus a gewsys þ̄r deg,
 “Judas, ow ry te a vynn,
 dre dha vay a reth mar wheg
 dhe neb a’ m torment þ̄r dynn.
 Mollath den, ha gour ha gwreg
 a dheu þ̄oran er dha bynn.
 Painys a’ th wra morethek
 in iffarn down̄ pub termyn.”*

Jesus spoke very fairly,
 “Judas, you will betray me,
 by your kiss, which you give me so sweetly,
 to him who will torment me very sharply.
 The curse of men, both husband and wife,
 will come exactly against you,
 Pains will render you wretched
 in deep hell forever.”

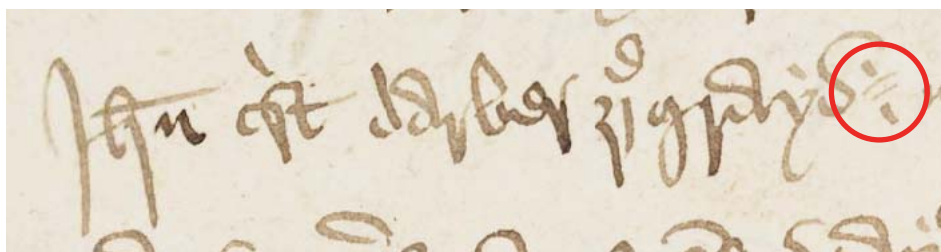


Figure 5. Text from *Beunans Meriasek* (Peniarth MS 105B, fol. 20r), showing DOUBLE OBLIQUE HYPHEN WITH FALLING DOTS. The text reads:

Ihū c̄ft darber ʒ̄j grays †
Jesu Cryst, darbar dha ras
 ‘Jesus Christ, provide thy grace’

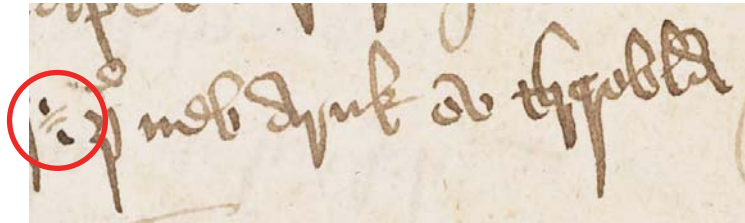


Figure 6. Text from *Beunans Meriasek* (Peniarth MS 105B, fol. 19r), showing DOUBLE OBLIQUE HYPHEN WITH FALLING DOTS. The text reads:

÷ ʒ̄ neb a ruk ov throbla
dhe neb a wrug ow throbla
 'For him who has troubled me'

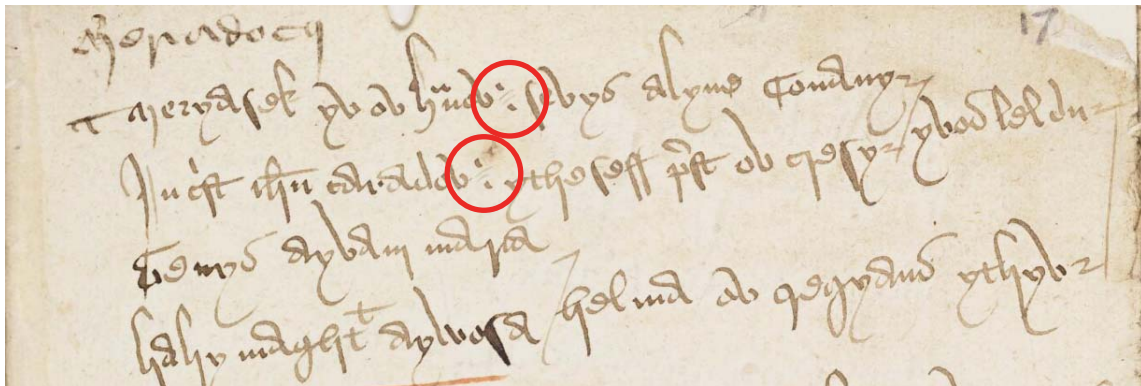


Figure 7. Text from *Beunans Meriasek* (Peniarth MS 105B, fol. 17r), showing DOUBLE OBLIQUE HYPHEN WITH FALLING DOTS. The text reads:

Meradocg
 Mezyafek yv ov hñov ÷ fevys a lyne Conany
 In c̄st ih̄u caradov ÷ ythefeff p̄st ov crefy / y vos lei du •
 Genys ay vam maria
 ha hy maght^t aywofa / helma ov cregyans yth yv •

MERIADOCUS
Meryasek yw ow hanow,
sevys a lin Conany.
Yn Cryst Jesu caradow
yth esof prest ow cresy.
Y vos lel Dew
genys a'y vam Marya,
ha hy maghteth a'y wosa;
hemma ow crejyans yth yw.

MERIASEK
 Meriasek is my name,
 Raised from Conan's line
 In Christ Jesu the loveable
 Am I ever believing.
 That he is true God
 Born of his Mother Mary,
 And she a maiden notwithstanding ;
 This is my belief.

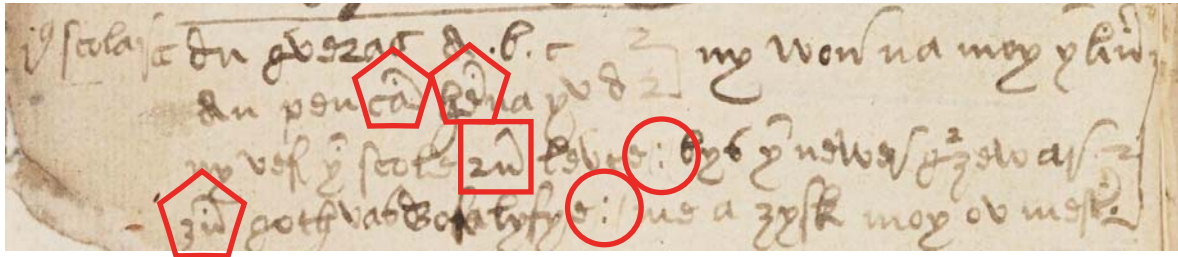


Figure 8. Text from *Beunans Meriasek* (Peniarth MS 105B, fol. 3v), showing DOUBLE OBLIQUE HYPHEN WITH FALLING DOTS. In this rather unskilful hand the dots and lines are reversed and appear like |:| or /:/ but the function is clear and unambiguous and this glyph is not to be preferred (the glyph could be composed with VERTICAL BAR and COLON at need). There is also an example of the COMBINING FINAL SWASH in this here (in *zu*, where it means *m*), alongside the COMBINING INVERTED BREVE (in *y'*). There are also three examples of a swash with a dot (one word-internally, so it may be that the name COMBINING SWASH may be considered) on *a* and *e* (where it means *n*) and *u* (where it means *m*), but in Cornish texts these are rare and it is recommended to use the sequence base + COMBINING SWASH + COMBINING DOT ABOVE; no *COMBINING SWASH WITH DOT is proposed. The text reads:

i' scolar
 du gvezas A · b · c / an pen can henna yv d
 ny won na moy y' liu'
 ny vef y' fcole zu levte z bys y' newer g'zewar
 3u gothvas wofa lyfye z me a zyfk moy ov mest'

PRIMUS SCOLARIS
Dew gweres A B C, -
an pen can, henna yv D.
Ny won na moy y'n lyver.
Ny vef yn scol, re'm leouta,
bys yn nyhewer gordhewer.
Dhe'm godhwos, wosa lyvya
me a dhysk moy, ow mester.

FIRST SCHOLAR
 God keep A, B, C,
 The end of the song, that is D.
 I know no more in the book.
 I was not at school, by my loyalty,
 Until late (?) yesterday evening.
 To my knowledge, after dining
 I will learn more, my master.

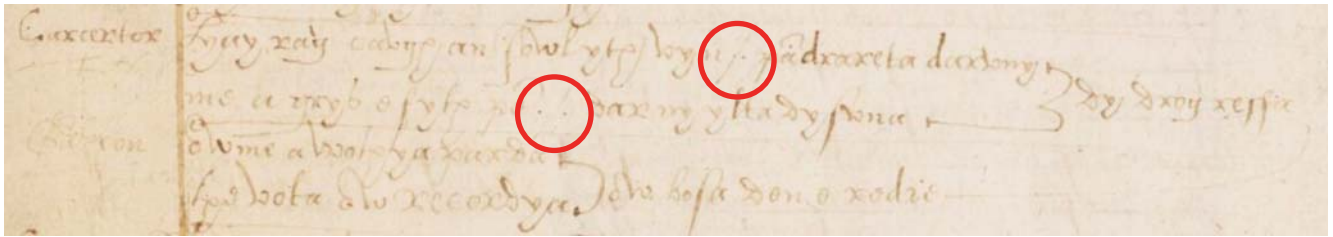


Figure 9. Text from *Bewnans Ke* (NLW MS 23489B, fol. 3v), showing DOUBLE OBLIQUE HYPHEN WITH FALLING DOTS. It was written around 1500 but survives only in an incomplete manuscript from the second half of the 16th century. The hand is spidery and untrained. Here, toward the beginning of the copy, the scribe draws the line separator horizontally, rather than at an angle, looking more like two middle dots and solidus \cdot/\cdot than as the paradigmatic character. See the next figure for the scribe's other glyph variants of the same character. The text reads:

Carcertor

Hay rag cavgh an Iowl yth vyn ꝛ pâdrareta darvny
 me a grys e fyth pê ꝛ dar ny ylta dyfvna
 dy drog reffa

CARCERATOR

Hay, 'rag! Caught an Jowl y'th vyn!
 Pandra reta? Dar, hunna,
 me a grys. E 'fyth ge peyn!
 Dar, ny ylta dyfuna?
 Dys! Drog re 'fo!

JAILER

Hey, come on! The Devil's turd in your mouth!
 What are you doing? What, sleeping,
 I think. You'll suffer!
 What, can't you wake up?
 Come on! Damn you!

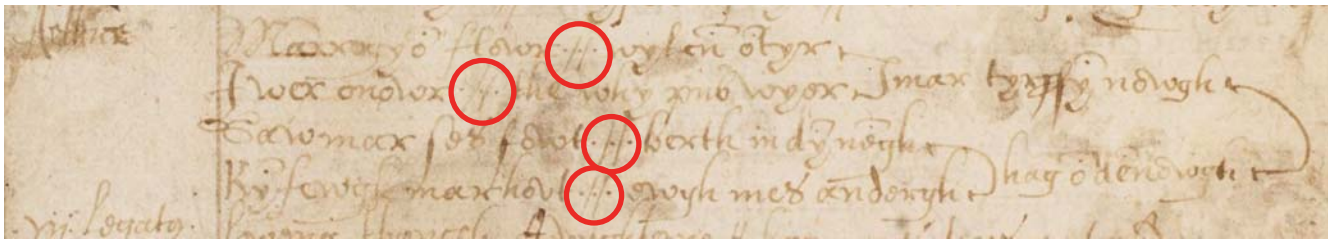


Figure 10. Text from *Beunans Meriasek* (Peniarth MS 105B, fol. 12v), showing DOUBLE OBLIQUE HYPHEN WITH FALLING DOTS. Here, much later in the copy, the scribe draws the line separator horizontally, with a variety of combinations of middle dots and solidus: ·/·. A form with two dots also occurs on this same page. ·//· The text reads:

Arthur

Marrogyô flowr ð wylcû ôtyr
 I wer onowr ð thewhy pub wyer
 mar tyrffÿnowgh
 Saw mar fes fowt ð berth in dÿnêgh
 Kÿfewgh mar hovt ð ewgh mes an dergh
 hag ôdênowgh

ARTHURUS

*Marrogyon flowr,
 wylcum o'm tyr,
 i wer onowr
 thewhy pub wyer
 mar tyrffÿnowgh.
 Saw mars es fowt
 aberth ynneugh,
 kyn fewgh mar hout,
 ewgh mes a'm veu
 hag omdennowgh.*

ARTHUR

Choice knights,
 welcome to my country,
 to great honour
 for you in very deed
 if you deserve.
 But if there is a fault
 within you,
 though you be never so haughty,
 get out of my sight
 and withdraw.

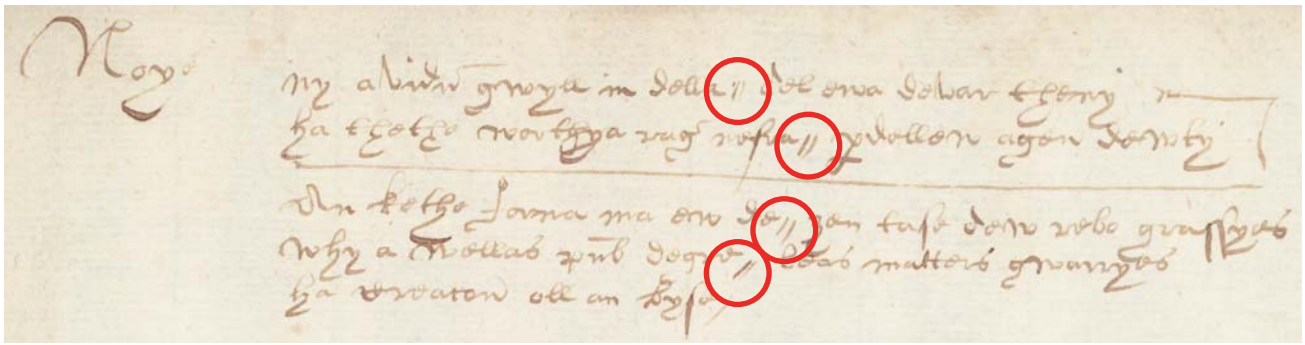


Figure 11. Text from *Gwryans an Bys* ‘The Creation of the World’ (Peniarth MS 105B, fol. 3v), showing DOUBLE OBLIQUE HYPHEN WITH FALLING DOTS. The scribe, William Jordan, does not run verse lines together very much but when he does he omits the dots. This document was written in 1611, a long time later than *Pascon agan Arluth*, so the scribal tradition established in Glasney has attenuated, as it has with *Bewnans Ke*. The text reads:

Noy

Ny a vydn gwyll in della Ꞥ del ewa dewar thenny,
 ha thethe wordhya rag nefra Ꞥ p dellew agan dewty
 An kethe jorna ma ew de Ꞥ 3en tafe dew rebo graffyes.
 Why a wellas pub degre Ꞥ leas matters gwarryes
 ha creacon oll an byfe

NOAH

*Ny a vydn gwil indella,
 dell ywa devar dhyn ny,
 ha dha wordhya rag nefra,
 par dell yw agan dũty.*

*An keth jorna-ma yw dũ,
 dhe'n Tas Duw re bo grassys.
 Why a welas pub degre,
 lies mater gwaries
 ha creacyon oll an bys.*

NOAH

We will do so,
 as is our obligation,
 and worship you forever
 as is our duty.

This same day is over,
 thanks be to God the Father.
 Every estate of you saw,
 many matters acted
 and the creation of all the world.

A. Administrative

1. Title

Proposal to add two characters for Medieval Cornish to the UCS

2. Requester's name

Michael Everson

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

Individual contribution.

4. Submission date

2017-09-27

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

2.

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 A.

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.

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?

The author is a member of the user community; also Nicholas Williams, Alan M. Kent.

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, Celticists, 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 and combining 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?

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

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?