

has some lacunae, it consists of over 20,000 lines of verse; the metrical nature of the work also assists in our understanding of the phonology of this dialect of Middle English. The edited text was published in White and Holt 1878, but their transcription partially normalizes Orm’s own scribal conventions and prevents the more comprehensive analysis of his orthography that a palaeographic reading can provide. The manuscript also contains numerous passages in Latin, and the scribe (who was Orm himself) used two distinctive styles of writing for each. When writing Latin, Orm uses a Carolingian hand with numerous signs of abbreviation and with a number of standard ligatures. He does not use these ligatures in his Middle English text, which is written in Insular script and is quite distinct from the standard Latin hand and orthography. It is for this reason that the palaeography of Orm’s orthography is both interesting and important. A palaeographic edition is being prepared, and in the preparation a number of characters missing from the UCS have been identified, The missing characters are proposed for encoding here.

1. COMBINING TRIPLE ACUTE ACCENT. Orm’s orthography marks vowels with three kinds of diacritics, drawn as a slightly angled straight line, two of those stacked, and three of those stacked. In modern orthography it is quite convenient to unify the first of these with U+0301 COMBINING ACUTE ACCENT and U+030B COMBINING DOUBLE ACUTE ACCENT; the third one is proposed for U+1ABF COMBINING TRIPLE ACUTE ACCENT. The specific meanings of all three of these accents is (as usual in Old and Middle English manuscripts) slightly uncertain, but they are clearly distinct and the missing one needs to be added to the UCS. Dickens and Wilson (1952:82) have suggested that readers trained in Latin might assume a short vowel before a single final *-t*, and that Orm’s accents were a reminder to pronounce. A comprehensive study of the distribution of these accents in the *Ormulum* has not been completed; certainly encoding the COMBINING TRIPLE ACUTE ACCENT will help such study. A similar set of spacing characters exists in the UCS: U+2032 ' PRIME, U+2033 " DOUBLE PRIME, and U+2034 ''' TRIPLE PRIME. It should be noted that single and double acutes in early English manuscripts are drawn with a somewhat less intense slope than the modern ACUTE ACCENT is, but the unification with the modern character is conventional, though in the examples in Figures 1 and 2 below they appear stacked rather than side-by-side; they are a bit more accent-like in the actual manuscript as shown in Figure 3. SC2 and the UTC should not encode “semi-sloped stacking macrons” for this character (they are distinct from either COMBINING MACRON and COMBINING OVERLINE).

á ă ǎ é ẽ ẽ ó ǒ ǒ

2. COMBINING INSULAR G, INSULAR R, AND INSULAR T. Orm’s orthography marks short vowels in two ways: by following the vowel with two consonants, and by following the vowel with a consonant topped with a smaller identical consonant. This sort of combining-character convention is normal in the medieval period. Compare two Middle Cornish spellings for “Crist” ‘Christ’:

crift ċft

In a few some cases the stack is replaced by a full “double letter” (see §4 below). These are not standard fancy-text ligatures, however: they are as distinctive in terms of Orm’s orthography as the letters surmounted by smaller letters, and in Orm’s English orthography he does not use any of the standard ligatures which he uses when he writes Latin. In fact, when the COMBINING OVERLINE is used to abbreviate *m* and *n*, he even doubles that to indicate a short vowel (so \bar{e} is *emm* or *enn*). In terms of the UCS, most of the letters can be encoded with a base letter and an existing combining character above.

ċ ḋ ħ ḣ ñ ṁ ñ ā ē ī ō ū

In Orm’s orthography. DOUBLE THORN and DOUBLE WYNN are typically used following a vowel, as doubled consonants are used to indicate vowel length. Orm is scrupulous about capitalization; indeed, his casing of TIRONIAN ET is regular and helped to get it encoded. It has not been possible for me to examine all 110,000 words in the 20,000 lines of verse in the *Ormulum*. It is certainly possible that Orm *didn’t* write any words in all-capitals. But if he did, it’s likely that he would have managed both CAPITAL DOUBLE THORN and CAPITAL DOUBLE WYNN. He knew what capital letters are, and how to use them.

But Orm is long dead. Orm doesn’t care about using Unicode or ISO/IEC 10646. It is modern scholars like myself who wish to produce palaeographic or expanded editions of the *Ormulum*, to cite words and phrases from the *Ormulum*, and indeed to use them in titles as given in the completely likely examples given above.

Latin letters used in natural orthographies are naturally casing. This is a standard structural function of the Latin script. We have encoded many casing pairs for natural orthographies in the past and we should not have to have the same argument every time new characters are proposed for encoding.

7. Glyphs. A variety of glyphs can be used to represent the INSULAR G. Some look z-like, some 3-like, some have a closed circular counter \mathfrak{g} , and some—the most iconic and common—look more S-like, without a closed counter. Orm uses the S-like glyph for both his /j/ and /g/ letters. A glyph for his /g/ having two closed counters does not look right. For consistency in the UCS, we recommend that the following glyphs be used for A77D, 1D79, A77E, A77F, A7D0, and A7D1:



8. Ordering. We recommend the following.

... g << G << ... << \mathfrak{g} << \mathfrak{g} << \mathfrak{g} << \mathfrak{g} << g << G ...

... r << R << ... << \mathfrak{r} << \mathfrak{r} << \mathfrak{r} << r << R ...

... t << T << ... << \mathfrak{t} << \mathfrak{t} << \mathfrak{t} << t << T ...

... þ << Þ << $\mathfrak{þ}$ << $\mathfrak{þ}$ << þ << Þ << þ << Þ << þ << þ << $\mathfrak{þ}$ << $\mathfrak{þ}$ << þ << Þ ...

9. Security. None of these characters are required in identifiers.

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

```
1AC9;COMBINING TRIPLE ACUTE ACCENT;Mn;220;NSM;;;;;N;;;;;
1ACA;COMBINING LATIN SMALL LETTER INSULAR G;Mn;220;NSM;;;;;N;;;;;
1ACB;COMBINING LATIN SMALL LETTER INSULAR R;Mn;220;NSM;;;;;N;;;;;
1ACC;COMBINING LATIN SMALL LETTER INSULAR T;Mn;220;NSM;;;;;N;;;;;
A7D0;LATIN CAPITAL LETTER CLOSED INSULAR G;Ll;0;L;;;;;N;;;A7D1;
A7D1;LATIN SMALL LETTER CLOSED INSULAR G;Ll;0;L;;;;;N;;;A7D0;A7D0
A7D2;LATIN CAPITAL LETTER DOUBLE THORN;Ll;0;L;;;;;N;;;A7D3;
A7D3;LATIN SMALL LETTER DOUBLE THORN;Ll;0;L;;;;;N;;;A7D2;A7D2
A7D4;LATIN CAPITAL LETTER DOUBLE WYNN;Ll;0;L;;;;;N;;;A7D5;
A7D5;LATIN SMALL LETTER DOUBLE WYNN;Ll;0;L;;;;;N;;;A7D4;A7D4
```

11. Bibliography

Dickins, Bruce, & R. M. Wilson, eds. 1952. *Early Middle English texts*. Cambridge: Bowes & Bowes.

- Hall, Joseph. 1920. *Selections from Early Middle English, 1130–1250*. Part I: Texts. Oxford: Clarendon Press.
- Napier, Arthur S. 1894. “Notes on the orthography of the Ormulum”, in *History of the Holy Rood-tree, a twelfth century version of the Cross-legend*. (Early English Text Society: Original Series; 103.) London: Kegan Paul, Trench, Trübner & Co. Pp. 71–74. Reprinted 1998 ISBN 0-85991-857-2.
- White, R. M., and Robert Holt, eds. 1878. *The Ormulum, with the notes and glossary of Dr R. M. White*. 2 volumes. Oxford: Clarendon Press.

12. Figures.

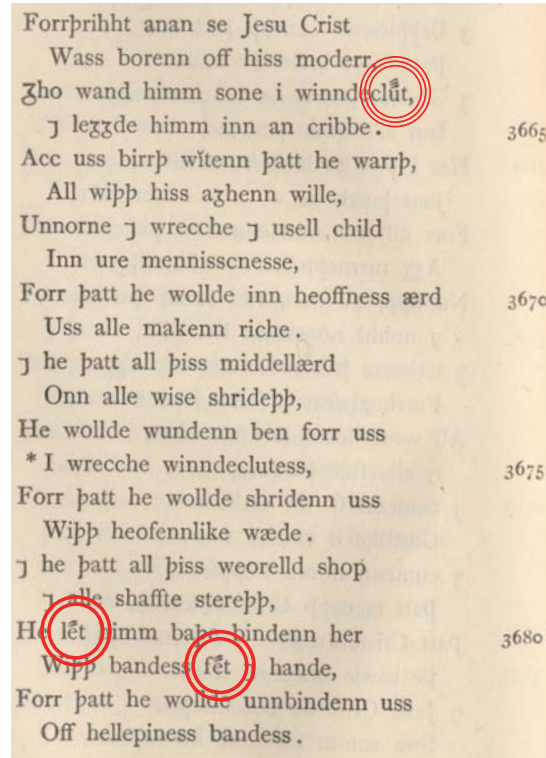


Figure 1. Example from White and Holt 1878:xcvii showing ◌̇̇̇ COMBINING TRIPLE ACUTE ACCENT alongside U+030B ◌̇̇ COMBINING DOUBLE ACUTE ACCENT.

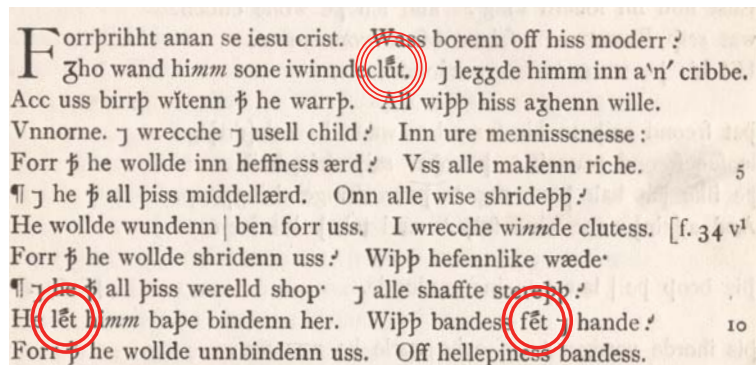
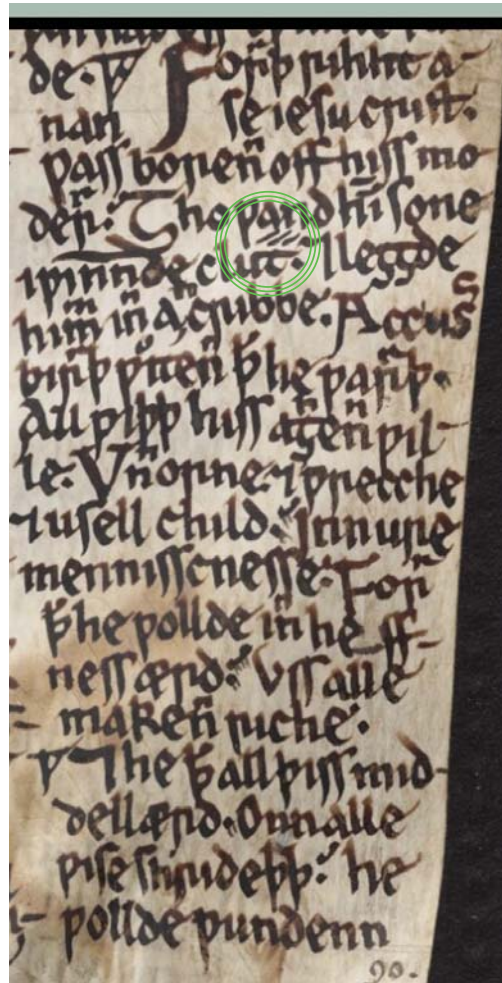
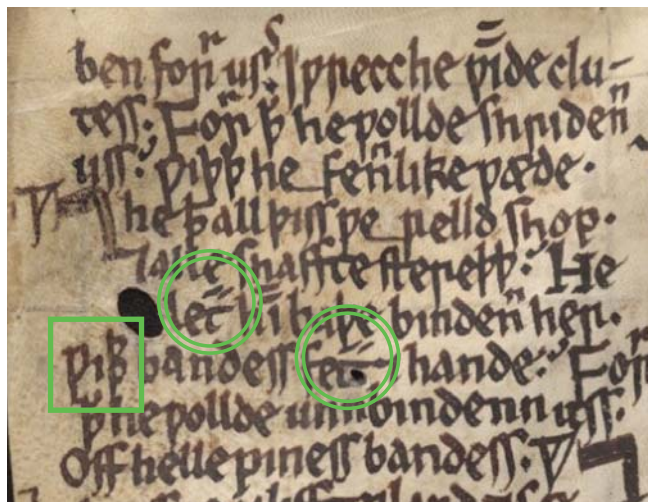


Figure 2. Example of the same text from Hall 1920:112 showing ◌̇̇̇ COMBINING TRIPLE ACUTE ACCENT alongside U+030B ◌̇̇ COMBINING DOUBLE ACUTE ACCENT.



f. 34r



f. 34v

Figure 3. The same text as given in Figures 1 and 2, from the *Ormulum* f. 34r–34v, lines 3662–3683, showing ◌̊̊̊ COMBINING TRIPLE ACUTE ACCENT alongside U+030B ◌̊̊ COMBINING DOUBLE ACUTE ACCENT. Sometimes for reasons of space it appears that Orm writes the accents atop the following character, but it is conventional (and proper) to read them as belonging to the vowel, given the design and intent of his orthography. Note in the very first line of the image from f. 34v the word *p̄de winnde* ‘swaddling’. Here U+0305 COMBINING OVERLINE indicating an *m* or *n* is used twice (because Orm doubles these too to indicate vowel length); this is not the COMBINING DOUBLE ACUTE ACCENT. There is also a SMALL DOUBLE THORN here, shown in the image in a box.

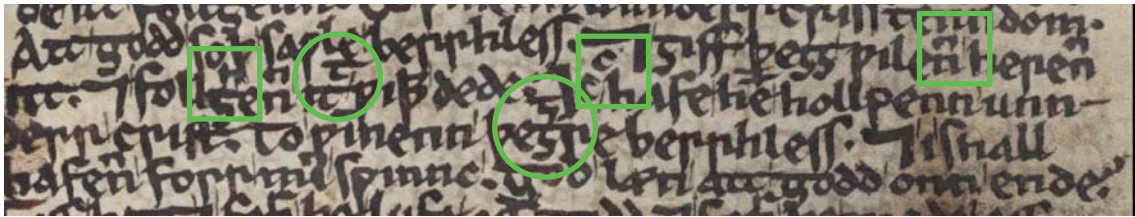


Figure 4. Example from the *Ormulum* f. 3v showing ◌̇ COMBINING LATIN SMALL LETTER INSULAR LETTER G in þeḡne (line 3) and ◌̇ COMBINING LATIN SMALL LETTER INSULAR LETTER T in iē (line 2), alongside examples of COMBINING C in Iē (line 2), of COMBINING H in follḡeḡ (line 2), and of COMBINING N in pileḡ (line 1). The four lines of text here read, in palaeographic transcription:

Att godd foḡ faḡle beḡḡheff. ⁊ ḡiff þeḡḡ pileḡ heḡeḡ
 itt. ⁊ follḡeḡ iē ḡiḡ dede. Iē hafe hē hollpenn unn-
 deḡḡ cḡiḡt. To ḡineḡḡ þeḡḡe beḡḡheff. ⁊ i fhall
 hafeḡ foḡḡ ḡiḡ ḡpinnc. Soḡ læn att godd onn ende:

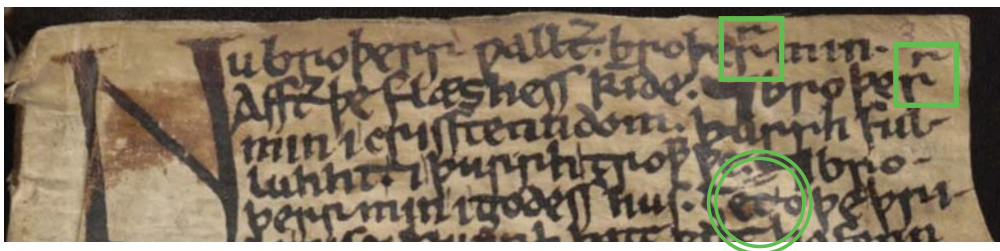


Figure 5. Example from the *Ormulum* f. 3v showing ◌̇ COMBINING LATIN SMALL LETTER INSULAR LETTER R in broḡeḡ *broḡerr* ‘brother’ (lines 1 and 2; broḡeḡḡ is written hyphenated between lines 4–5). The COMBINING DOUBLE ACUTE ACCENT can be seen in line 5. The five lines of text here read, in palaeographic transcription:

Nu broḡeḡḡ ḡallē. broḡeḡḡ ḡin.
 Affē þe flæsheff kīde. ⁊ broḡeḡḡ
 ḡin i cḡiḡḡtenndom. Þurḡḡ ful-
 luhht ⁊ þurḡḡ tḡoḡḡe. ⁊ bro-
 þeḡḡḡ ḡin i ḡodeff huf. (Sēt) o þe þḡi-

For the guttural and palatal spirants Orm retains the Old English γ (in the former case adding an h), while for the guttural stopped consonant, as in *god*, he employs a sign which is represented in the printed editions by g . But in the printed editions the sign g is not restricted to the guttural stopped consonant: it is used also in words like *egge* (= Modern English 'edge'), where the gg had the dzh sound. In other words, the printed editions of the Ormulum make no difference between *egge* (= 'edge') and *eggenn*² (= to 'egg on'), though the pronunciation of the consonants in the two words was, in Orm's time, the same as now, i.e. dzh in the former case and a stopped g in the latter. But on examining the MS. I found that, though the editors make no difference, Orm did. The letter with which he always denoted the guttural stop (as in *god*, *gladd*, *eggenn*, &c.) is perfectly distinct from the sign which he used to express the dzh sound (as in *egge* 'edge,' *seggen* 'to say,' &c.). The latter, which in the following remarks, as well as in the transcript of the facsimile, is denoted by g , has the form of the continental g —: cf. *biggenn* in the facsimile, lines 2, 16, 18, 20, 41, 47, and *seggenn*, line 6. The former, which I shall denote by \mathfrak{g} , may be described as a sort of compromise between the Old English γ and the continental g : it has, in common with this latter, the closed upper part, thus differing from the Old English γ ; but it has, in common with the Old English γ , the straight horizontal top stroke, which projects to the left as well as to the right of the letter —: cf. *godd*, lines 4, 9, 10, 12, 38, &c., *biginnen*, line 13, &c. This straight horizontal top, especially that part of it which projects to the left, is its most characteristic feature, and serves to distinguish it from the g , from the round top of which a short sloping stroke extends to the right, there being no stroke whatever to the left. The absence of any stroke to the left of the top of the g at once distinguishes it from the \mathfrak{g} . Except for the one or two isolated instances mentioned on page 4, Orm never confuses the two signs, but always uses them correctly, \mathfrak{g} denoting the guttural stopped consonant, and g the dzh sound. I give a few instances—the pronunciation, g or dzh , is added in brackets, the number which follows denotes the number of times I have met with the word in question in the Ormulum MS.: *egge* 'edge' (dzh -4) is in each case written with gg ; *eggenn* 'to egg on' (g -5), *egginn* \mathfrak{g} (g -1) are in every instance spelt with \mathfrak{gg} . The verb *biggenn* 'to buy' (dzh -18) is always written with gg , being thus invariably distinguished from *biggenn* 'to dwell' (g -20). The verbs *leggenn* 'to luy' (dzh -2), and *seggenn* 'to say' (dzh -33) are in every instance written with \mathfrak{gg} , while the Scandinavian *tri \mathfrak{g}* 'faithful' (g -3), *kaygerrlez \mathfrak{g}* 'love' (g -2) are spelt with \mathfrak{gg} .

Figure 6. Discussion in Napier (1894:71–72) of his analysis of the distinction Orm makes between his invented \mathfrak{g} / g /, \mathfrak{g} / j /, and g / $dʒ$ /. The graphs \mathfrak{gh} and \mathfrak{h} were used for / y /. The glyphs used in this Early English Text Society publication are not the best, though they do the job. (Note that an italic form of g was not cut.) But the shape of the insular G and the closed INSULAR G should be more similar.

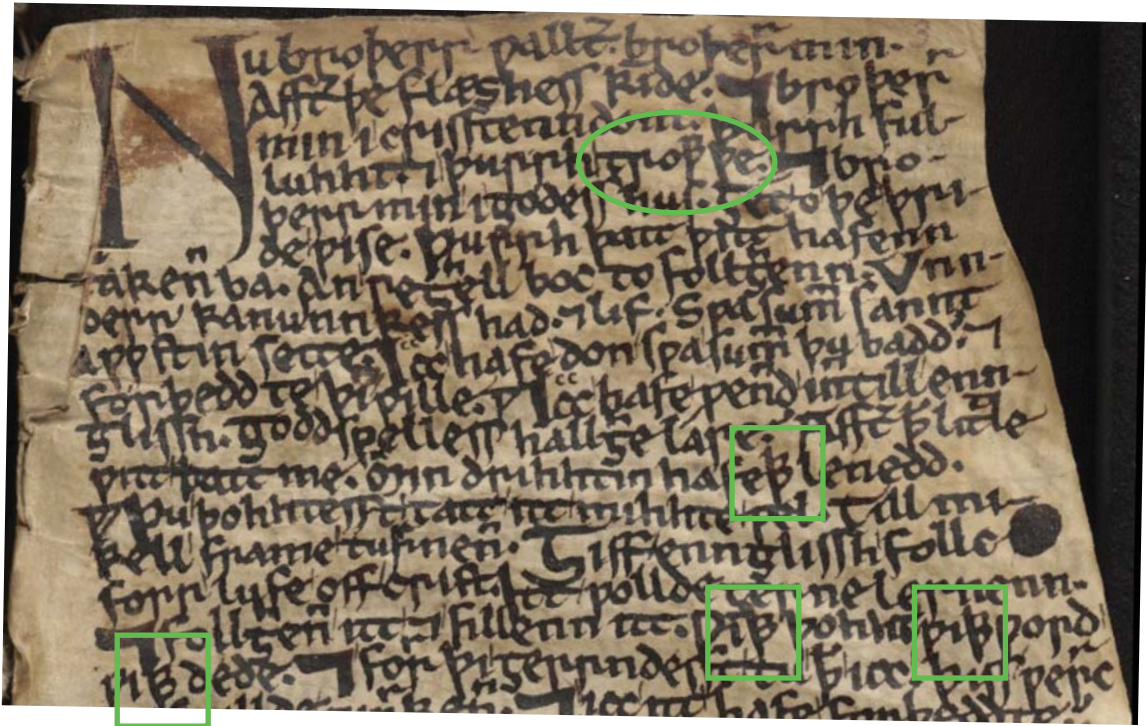


Figure 8. Example from the *Ormulum* f. 3v showing LATIN SMALL LETTER DOUBLE WYNN in the word *trowþe* (*trowþe* ‘belief’) and LATIN SMALL LETTER DOUBLE THORN in the words *hafaþ* (*hafaþþ* ‘has’) and *wiþ* (*wiþþ* ‘with’). The text here reads, in palaeographic transcription:

Nu broþerþ Pallè. broþerþ min.	f.3r	Y Icē hafe peñd iñtill enngliſh.	
Affè þe flæsheff kīde:		Soddselleſſ hallþe lápe:	
7 broþerþ min i cwiſſtenndom.		Affè þ liçle piçt þatt me.	15
Þurþ fulluhht 7 þurþ trowþe :		Min drihhtin hafaþ lenedd.	
7 broþerþ min i godeſſ huſ.	5	Y Þu þohhteſſt tatt itt mihhte pel.	
Œt o þe þpide piſe.		Cill mikell frame tuþneþ,	
Þurþ þatt piçt hafenn takeñ ba.		Siff enngliſh folc forþ luſe off cwiſt.	
An reþellboc to follþenn.		Itt pollde ſerne leþnenn.	20
Vnnderþ kanunnkeſſ had. 7 lif.		7 follþeñ itt 7 fillenn itt.	
Spa fuþm fannt Appſcin ſette:	10	Þiþ þohht, þiþ þorþd, þiþ dede.	
Icē hafe don ſpa fuþm þu badd:			
7 forþpedd te þi pille.			

A. Administrative

1. Title

Revised proposal to add ten phonetic characters for Scots to the UCS

2. Requester's name

Michael Everson and Andrew West

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

Individual contribution.

4. Submission date

2020-10-05

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

Combining Diacritical Marks Extended (4); Latin Extended-D (6)

2. Number of characters in proposal

10.

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. 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.

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.)?

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?

Germanicists, Anglicists, dialectologists, linguists.

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

Used in the *Ormulum*, a unique but very important record of Early Middle English; also used in publications about it and extracts from it.

4b. Reference

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

Yes.

5b. If YES, where?

Various 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 Latin 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?

No.

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

Cross references point to the related but different 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)?

Yes.

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

No.

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?