Doc Type: Working Group Document
Title: Preliminary proposal to add medievalist characters to the UCS
Source: Michael Everson, Odd Einar Haugen, António Emiliano, Susana Pedro, Florian Grammel, Peter Baker, Andreas Stötzner, Marcus Dohnicht, Diana Luft
Status: Expert Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2005-08-02

Introduction.
A set of characters used by specialists in medieval European philology and linguistics is absent from the Universal Character Set. These characters differ in nature; some are original ligatures which acquired letter status due to their phonemic value; some are letterforms distinct from other letterforms innovated to distinguish sounds; some are combining diacritical letters used in abbreviations or suspensions of various kinds; and some are best described as "letters with syllabic content". *This is not a complete proposal, but a work in progress; all of the example Figures referred to will be added in the next version of this document. Note that the glyphs used in the code charts here may not be optimal.*

Theoretical preliminaries.
Contemporary medievalist philologists and linguists want to be able to represent typographically (in printed format and on computer screens) the *character sets* which were in use for many centuries in several regions of medieval Europe. Those character sets derive from the common Latin script and contained many characters which simply disappeared with the development of contemporary printing conventions. Early printers made abundant use of "special" medieval characters, but eventually these fell out of use, with notable exceptions like $, ⸿, &, Ç, ˜, @, and the ⁊ used in Ireland.

Contemporary philologists and linguists who want to study the graphemic conventions in use in medieval times—thereby drawing solid or grounded conclusions about the nature and structure of the language systems represented in writing—must rely on *bona fide transcriptions* of the texts. Bona fide transcriptions are only possible when the elemental character set used in the manuscripts is encoded uniquely and available for use in fonts.

What most philologists did in the 20th century was to publish *transliterations*, that is, editions which substitute modern characters (or sequences of modern characters) for the original medieval characters. Transliteration-based editions are virtually useless for those scholars who are interested in the study of medieval writing systems, phonology, and even textual structure. Transliterations (or "normalized editions") and even translations may, of course, be required for editions aimed at students or the general public, but the base texts must result from *transcribing* the sources: the first step must always be a transcription.

Transcription is an editorial process which does not entail the replacement or the distortion of the original character set. A bona fide transcription of a Runic, Coptic, or Egyptian Hieroglyphic text can be no less than a close rendition of the original characters, using typographic versions of Runic letters, Coptic letters, and Hieroglyphic characters. The same practice must apply to European medieval texts.

The practice of *expanding abbreviations* common to medievalists throughout the 20th century is not *transcription*, but simply *transliteration*: abbreviations, or brachygraphemes, were special characters. (Brachygraphy is, according to the *Oxford English Dictionary*, "The art or practice of writing with abbreviations or with abbreviated characters; shorthand, stenography.") Many of these graphemes were

polyvalent—that is, they could be transliterated into different sequences of "normal letters", according to textual context, country, region, time period, and even individual scribal practices. Polyvalence is not a uncommon feature of alphabetic writing systems, as our own modern spelling systems show; for instance, in European Portuguese orthography the letter E can have the values [ɛ], [e], [ɨ], [i], [j], [i̯], [ĩ], and Ø; this letter combined with *-m* or *-n* can further represent [ẽ] and [ĕĩ]. Graphemic polyvalence results in many instances from language change and from the conservative nature of spelling systems.

Another common practice in the 20th century was to eliminate the original punctuation and to add modern punctuation—many scholars believed that medieval punctuation served no discernable purpose. "Modernization" of the use of capital letters was also standard practice.

A medieval character set makes it possible to shed many "chronocentric" biases and prejudices which tainted many editorial efforts in the 19th and 20th centuries in many countries, rendering the ensuing editions virtually useless for at least some kinds of contemporary research. This *does not mean* that medievalist scholars wish or even need to represent in print every minutia that handwritten sources present—that is palaeography proper. In medieval texts this is a particularly delicate issue, because scholars have to deal with a considerable amount of regional or individual stylistic variation. The rationale behind encoding medieval characters and designing medieval fonts is not to capture in print every single glyph variation (a task which is virtually impossible and also meaningless), but to capture the *character set* used in the manuscripts under scrutiny. We understand the character/glyph model and how it applies to the medieval character set.

Accurate transcriptions of medieval texts allow scholars to quote medieval texts without distorting their graphemic content, and allow the texts to be studied by means of computer applications such as concordance generators and wordlist generators. Accurate transcriptions which make use of a medieval character set are a means to preserve—and interchange—all the relevant graphemic, textual, and linguistic information contained in a text; they are also indirectly a means to contribute to the preservation of Europe's early heritage.

Case-pairing.
Most of the casing pairs shown below are attested in the examples. Those which are not, fall into two categories: those for which no capital can be constructed (such as LONG S) and those for which natural capitals can be easily formed. In an early version of this document we had proposed a single lower-case character <ɏ> LATIN SMALL LETTER Y WITH STROKE in use by some Welsh medievalists to indicate an epenthetic schwa sound (Figure 8). Subsequently we discovered that this character *and its capital* are ballotting in FPDAM2, as a character used in the Lubuagan Kalinga language of the Philippines. Because of the general structural feature of the Latin script (from a theoretical point of view), and in order to facilitate modern casing operations for these letters, we have judged it appropriate to supply case-pairs for all the letters which admit of them. In a scholarly publication, for instance, an article title at the top of a journal page might be set in all caps; it would be nonsensical for all but one or two of the medievalist Latin letters to be able to be cased with an all caps command. (This precedent was set with the encoding of the archaic Coptic extensions.)

Discussion.
1. Letters used for medieval Welsh.

<ɪʟ ꜹ>   The voiceless lateral fricative written <ll> in modern Welsh may be written in medieval Welsh as a joined ligature of <ꜹ> LATIN SMALL LETTER MIDDLE-WELSH LL. Its capital form <ɪʟ> LATIN CAPITAL LETTER MIDDLE-WELSH LL has been used as an abbreviation in Welsh scholarship (Figures 1, 2, 3, 4, 5, 52).

<ð>   The voiced dental fricative written <dd> in modern Welsh is written <ð> LATIN SMALL LETTER INSULAR D distinctly from <d> in some editions of medieval Welsh texts; Nordic medievalists also make use of this letter (Figures 1, 5, 21, 29, 39, 40, 53, 73).

<δ>   Some Welsh medievalists (and other Indo-Europeanists of a certain era) also use <δ> LATIN SMALL LETTER SCRIPT D to write this sound in transcription. While this letter may sometimes have

been *represented* in print by using a Delta from a Greek lead-type font, it *derives* from the handwritten Latin d, and behaves like a Latin letter in ordering and is found *alongside* Greek text proper (Figures 6, 7, 8, 41).

<Ỽ ỽ>   A unique <ỽ> LATIN SMALL LETTER MIDDLE-WELSH V is used distinctly from <u>, <v>, and <w>, though it is true to say that the phonetic value of all four of these letters is polyvalent in medieval Welsh (Figures 3, 9, 10, 11, 12, 13).

<Ꝥ ꝥ>   Some Welsh medievalists use LATIN SMALL LETTER Y WITH LOOP to indicate the schwa sound of <y> (Figure 14).

<ꝛ ꝛ>   As in many medieval traditions, <ꝛ> LATIN SMALL LETTER R ROTUNDA is distinguished from <r> LATIN SMALL LETTER R. This named character is derived from a positional variant of <r> following <o> in the South Italian Beneventan, though in medieval Welsh and Nordic it is not limited to this position. In any case, Welsh and Nordic medievalists distinguish R from R ROTUNDA in their printed editions; the letter is also common in early printed texts throughout Europe (Figures 1, 3, 16, 17, 21, 24, 35, 36, 38, 42, 70, 71, 73). The case-pairing <ꝛ> LATIN CAPITAL LETTER R ROTUNDA is attested in printed texts from the 15th century (Figure <u>XXX</u>).

2. Letters used for medieval Nordic vowels.
Medieval Nordic orthographies innovated a number of letters out of original Latin script ligated letters. Some of these letters are well known today, as the letters <Æ æ Œ œ>. In medieval Nordic:

<Aa aa> LATIN LETTER AA is used for phonemic /aː/ (Figures 18, 19, 23, 28, 34).

<Ao ao> LATIN LETTER AO is used for phonemic /ǫ/ (Figures 15, 16, 17, 20, 66)

<Au au> LATIN LETTER AU is used for phonemic /au/, /ǫ/, /ø/, and /øː/ (Figure 38, 72)

<Av av> LATIN LETTER AV is used for phonemic /au/, /ǫ/, /ø/, and /øː/ (Figures 18, 21, 24, 42, 70, 72)

<Av̵ av̵> LATIN LETTER AV WITH HORIZONTAL BAR is used for phonemic /ǫ/, /ø/, /øː/, and /ę/ (Figure 24)

<Ay ay> LATIN LETTER AY is used for phonemic /ǫ/, /øː/, and /ey/ (Figure 69)

<Oo oo> LATIN LETTER OO is used for phonemic /oː/, (Figure 28)

<Oʼ oʼ> LATIN LETTER O WITH LOOP is used for phonemic /ǫ/, /øː/, and /ey/ (Figures 19, 25, 35, 36)

<Vy vy> LATIN LETTER VY is used for phonemic /yː/ (Figure 28). This collection of characters is a superset of the letters found in the medieval Nordic corpus; no single manuscript contains all of them. Note also that none of these characters is a "ligature" that can be broken; indeed, all of them are known to bear diacritical marks.

3. Letters used for medieval Nordic consonants.
<Ƚ ƚ>   LATIN LETTER BROKEN L was used for phonemic /lː/ (Figure 70)

<Ỿ ỿ>   LATIN LETTER VEND (ultimately derived from the English letter WYNN) for phonemic /v/ or /u/. Some editions use u, v, and ỿ in the same text (Figures 21, 26, 35, 36, 42).

<ꜰ s>   The Icelandic First Grammarian's orthography made use of small capital letters to indicate gemination of consonant sounds, as Uralic linguists did centuries later. Between letters encoded already for Uralicist and IPA use, most of the Latin alphabet is already encoded as small capitals; while <ꜰ> SMALL CAPITAL F and <s> SMALL CAPITAL S are yet missing from the UCS (Figure 22). It should be noted that of the traditional Latin alphabet, if these two are added, only *SMALL CAPITAL Q and *SMALL CAPITAL X will remain unencoded.

4. Letters used for medieval Ibero-Romance.
<ẜ ẜ>   The Latin alphabet we use today is only one of several variants. Our own lowercase "Roman" type is derived from the Carolingian variant of the Latin script; the Insular and Germanic variants are fairly familiar to us, having enjoyed a period of typographic development as Gaelic and Fraktur, and a handful of letters from the Insular tradition have been adopted by the Carolingian tradition for one purpose or another (WYNN (as VEND) was used in Old Icelandic and Old Norwegain until ca. 1300; THORN and ETH are still used in Icelandic; Insular g and d have

been resurrected by linguists). The Visigothic variant of the Latin script, however, was replaced before the advent of typography, and its unique letterforms were simply lost to the Carolingian, apart from the <ʓ ʓ> ʟᴀᴛɪɴ ʟᴇᴛᴛᴇʀ ᴠɪsɪɢᴏᴛʜɪᴄ ᴢ. The Carolingian script was introduced in Northern Iberia in the 11th century—in Catalonia, the Spanish Mark of the Carolingian Empire, it was introduced earlier—but it only gained widespread use in the course of the the 12th century; the Visigothic script was extinct in the second half of the 12th century (1172 is the date for the last known original Portuguese document). The Visigothic <ʓ> was employed alongside the Carolingian <z>, and came to be used mainly to represent the voiceless alveolar affricate [ts], while <z> was used mainly for the voiced alveolar affricate [dz] in Old Portuguese, Old Leonese, and Old Castilian. In time, as Carolingian practices replaced all memory of the Visigothic, the head of the <ʓ> was reanalyzed, its tail reduced, resulting in a new letter <ç>. While in modern analysis the tail <˛> is known as a *cedilla/zedilla* 'little z', in fact the whole letter is, in origin, a <z>. Documents exist in which <ç> and <ʓ> and <z> are distinct (see Figures 43, 44).

5. Other letters of the Insular tradition. One of the letters of the Insular tradition has already been encoded at U+1D79. (A large number of letters in the Fraktur tradition have been encoded for use in mathematics.) The set of Insular letters which differ significantly enough from Carolingian to warrant distinction is small; medievalists have used them in typeset editions of Germanic and Celtic languages since the 16th century. *Modern* Germanic and Celtic languages do not use these letters, and *modern* Germanic and Celtic fonts which use Fraktur and Insular letterforms employ them as glyph variants pertaining to the entire font. The Insular letters proposed here are only to facilitate the specific need of historical linguistic specialists to differentiate the Insular letters from the Carolingian.

Insular and Carolingian letters coexisted but were often used in different contexts in Britain and Ireland in the Early Middle Ages, for example Insular letters being used for writing English and Carolingian for Latin. They were also mixed to varying degrees, and this unique variant of the Latin alphabet was exported to the Nordic countries in the 11th century; due to its dual inheritance it has often been termed Carolingian-Insular. Since the letter shapes of Insular and Carolingian script ultimately derive from Uncial script, the majority of Insular and Carolingian letters are basically identical, but a handful letters had quite distinct shapes and usage in Insular script, as we know it from English and Nordic writings. Four of these letters are already in the Standard; ᴛʜᴏʀɴ (from the Runic alphabet), ᴇᴛʜ, ᴡʏɴɴ (also from the Runic alphabet), and ɪɴsᴜʟᴀʀ ɢ. The letters ᴛʜᴏʀɴ and ᴇᴛʜ are still used in Icelandic, while ᴡʏɴɴ was accepted by the Standard due to its usage in early English sources and ɪɴsᴜʟᴀʀ ɢ on foot of its usage as a phonetic character. We now propose to add five distinct letter forms to the Standard, i.e. ɪɴsᴜʟᴀʀ ᴅ, ɪɴsᴜʟᴀʀ ꜰ, ɪɴsᴜʟᴀʀ ʀ, ɪɴsᴜʟᴀʀ s, and ɪɴsᴜʟᴀʀ ᴛ. It should be underlined that it is not a question of adding Insular variants of every Latin character; it is a short list of distinctive letters that have been recognised as separate characters for several centuries in Medieval English and Nordic writing, and which have been used alongside and in contrast to their Carolingian counterparts. In Medieval Nordic editorial practice, these letters are rendered as separate characters in great many editions and distinguished from their Carolingian-based counterparts <d>, <f>, <r>, <s> and <t>. This is in part because the presence of these letters is used as a dating criterion (for example, ɪɴsᴜʟᴀʀ ʀ fell out of use around 1200, while ɪɴsᴜʟᴀʀ ꜰ continued to be used well into the 14th century), and in part because they are used in contrast to their Carolingian counterparts. In Ælfric's Old English grammar, the scribe, and the modern medievalist, distinguishes between <ꝼ> and <f>, between <ꞃ> and <r>, between <ꞅ> and <s>, and between <ꞇ> and <t> (Figure 39). In the sample from the edition of AM 645 4to <ꝼ> and <f> are distinguished (Figure 38).

<Ꝼ ꝼ>   ʟᴀᴛɪɴ ᴄᴀᴘɪᴛᴀʟ ʟᴇᴛᴛᴇʀ ɪɴsᴜʟᴀʀ ꜰ (Figure 70), ʟᴀᴛɪɴ sᴍᴀʟʟ ʟᴇᴛᴛᴇʀ ɪɴsᴜʟᴀʀ ꜰ (Figures 29, 30, 32, 33, 35, 36, 37, 38, 39, 40, 42, 70, 71, 73)
<ꞃ>   ʟᴀᴛɪɴ sᴍᴀʟʟ ʟᴇᴛᴛᴇʀ ɪɴsᴜʟᴀʀ ʀ (Figures 29, 30, 37, 39, 40)
<ꞅ>   ʟᴀᴛɪɴ sᴍᴀʟʟ ʟᴇᴛᴛᴇʀ ɪɴsᴜʟᴀʀ s (Figures 16, 29, 30, 37, 39, 40)

<τ>　　LATIN SMALL LETTER INSULAR T (Figures 29, 30, 37, 39, 40)

6. Letters used for medieval abbreviations. Medieval manuscripts, in both Latin and vernacular languages, use abbreviations extensively. Many of these are abbreviations for whole words, created by omitting letters, such as *sp̄s* for *spiritus*; often a line is placed over the letter(s) as an abbreviation marker, as shown here. Such "logographic" or "lexical" abbreviations can usually be represented through characters already encoded in the UCS. In other cases, however, only a part of a word is abbreviated; for example, the prefix *con-* is represented with the letter ɔ. A number of such syllabic abbreviations, well-documented and commonly used in several languages, require letters or combining marks that are not in the UCS. A range of Latin letters, modified by strokes or hooks, is used to represent a variety of words, syllables, or quasi-syllabic letter sequences. That they are polyvalent is a chief indicator for the requirement to encode these "abbreviation letters" as characters, since they cannot be composed of any specific string of other characters; neither can they be decomposed into a single string. Of these abbreviations:

<Ꝁ ꝁ>　LATIN LETTER K WITH STROKE is used for *karta, kartula, kalenda* (Figures 33, 45, 74)

<Ᵽ ᵽ>　LATIN LETTER P WITH STROKE THROUGH DESCENDER is used for Latin and Romance *per, par, por*, and for *pri* in Cornish *ᵽvecter priᵽvecter* 'privacy' (Figures 46, 47, 50, 58, 60, 63)

<ꝑ ꝑ>　LATIN LETTER P WITH FLOURISH is used for *pro, por* (Figures 46, 50, 52, 54, 55, 58, 60, 61, 63)

<Ꝗ ꝗ>　LATIN LETTER Q WITH STROKE THROUGH DESCENDER is used for *quam*, *que*, *quan-* (*ꝗdo quando*, *ꝗtum quantum*), *qui-* (*ꝗƺ quilibet*, *ꝗdem quidem*) and in Irish for *ar* (Figures 16, 17, 46, 49, 51, 53, 59, 63).

<Ꝁ ꝁ>　LATIN LETTER K WITH DIAGONAL STROKE is used for *kalendas* and *karta* (Figure 45)

<Ꝁ ꝁ>　LATIN LETTER K WITH STROKE AND DIAGONAL STROKE is used for *karta*(*m*) 'document, writ' when <ꝁ> LATIN SMALL LETTER K WITH STROKE is used for *karta* (Figures 45)

<Ƚ ƚ>　LATIN LETTER L WITH HIGH STROKE is used for Latin *el, ul, vel*, for Irish *nó* 'or', for Norse *eða* 'or', for *el* in *vƚ vel* 'well', for *æl* in *mƚti mælti* 'spoke', for *al* in *skƚ skal* 'shall' (Figure 32, 51, 52, 60, 73)

<Ꝋ ꝋ>　LATIN LETTER O WITH LONG STROKE OVERLAY is used for Latin *obiit* 'he died' (Figures XXX)

<ꝓ ꝓ>　LATIN LETTER P WITH SQUIRREL TAIL is used for Latin *prae* 'before, in front of' (Figures XXX)

<Ꝙ ꝙ>　LATIN LETTER Q WITH DIAGONAL STROKE is used for Latin *quod* 'what', *qui* 'who', *que* 'that', Portuguese *q̃ quem* 'who', Irish *ar* 'on' (Figures 58, 59, 60, 61, 63, 64)

<ſ>　LATIN SMALL LETTER LONG S WITH DIAGONAL STROKE is used for *ser* (*ſvir servir* 'to serve'), *sere* (*ſno sereno* 'serene'), *sir*, and by itself it also stands for Latin *solidi*, *sed*, *sunt*, *secundum*, etc., and for Portuguese *soldo*(*s*) (Figures 46, 48, 58)

<ſ>　LATIN SMALL LETTER LONG S WITH STROKE is used in Norse with <ƚ> for *ſƚ skal* 'shall' (Figures XXX)

<Ꝟ ꝟ>　LATIN LETTER V WITH DIAGONAL STROKE is used for Portuguese *ver* 'to see', *conꝟsa conversa* 'conversation', for *vere* in *ꝟador vereador* 'town councillor', *uirgo* 'virgin', and *vir* 'to come' (Figure 65)

<Ꝧ ꝧ>　LATIN LETTER THORN WITH STROKE is used for *þat, þess, þor-, þæt* (Figures 29, 32, 33, 73)

<Ꝥ þ>　LATIN LETTER THORN WITH STROKE THROUGH DESCENDER is used for *þeim, þeir* (Figures XXX).

7. Letters with syllabic content. This set of characters are also abbreviations, but might be better considered as "letters with syllabic content", because their reading tends to be less polyvalent than those of the abbreviation characters discussed above.

<ꝫ ꝫ>　LATIN SMALL LETTER ET is used for *-et* in *viꝫ videlicet* (whence "*viz.*"), *habꝫ habet* 'has', for *-m* in *ablůneꝫ ablutionem* 'ablution', for *-ue* in *usqꝫ usque* 'until, till, up to', for *-que* in *quiꝫꝫ quicumque* 'whoever', for *-us* in *ȧquibꝫ aliquibus* 'to someone', for *-est* in Latin *potest potꝫ* 'is

able', and for medial and final *eð* in Norse *mʒ með* 'with', *mʒan meðan* 'while' (Figures 58, 59, 60, 61, 63, 64)

<Ɔ ɔ>　　LATIN SMALL LETTER REVERSED C is used for *con* and *com* in *ɔⁱoɔ convenientior* 'more suitable, in agreement', *ɔpaᵗ comparavit* 'bought', for *-us* and *-os* in Latin in *totɔ totus* 'all', *tuɔ tuos* 'your', in Portuguese in *todɔ todos* 'all', *doɔ dous* 'two', and in Old Icelandic *ɔa kona* 'woman' (Figures 46, 50, 66). This is the same character as the Claudian letter discussed in §10 below.

<Ꙩ ꙩ>　　LATIN SMALL LETTER REVERSED C WITH DOT is used for *con-* and *com-* in *ꙩfmas confirmans* 'confirming, witnessing', *-us* and *-os* in *soldꙩ soldos* (for *solidos*) 'a unit of currency', *maladꙩ malados* 'serfs' (Figures XXX)

<ȣ ȣ>　　LATIN SMALL LETTER IS is used for *-is* in *dtȣ dictis* 'from having said', *impȣ imperatoris* 'ruler, emperor', and for *ys* and *es* in Cornish manuscripts: *godȣ godys* 'god's', *servantȣ servantes* 'servants', *mettȣ mettys* 'met' (Figure 57).

<ꝯ ꝯ>　　LATIN SMALL LETTER CON is used for *con* and *cum* and *co* (Figures 46, 47, 48, 53, 61)

<ꝰ>　　MODIFIER LETTER US is used for *-us* in *manꝰ manus* 'hand', *idꝰ idus* 'ides' (and thousands of other words) and final *os* in Latin, Portuguese, and Castilian: *oleyrꝰ oleyros* 'potters', *nꝰ nos* 'we', *uꝰ uos* 'you', and for *us* in Norse: *hꝰ hús* 'house' (Figures 50, 52, 59, 66).

<d̨>　　LATIN SMALL LETTER DUM is used by itself for Latin *dum* 'while, whilst', *die* 'day', Portuguese *dia* pl. *dias* 'day'

<m̨>　　LATIN SMALL LETTER MUM is used for *-mum* in *prim̨ primum* 'first' (Figures XXX)

<n̨>　　LATIN SMALL LETTER NUM is used for *-num* in *aetern̨ aeternum* 'eternal', *un̨ unum* 'one' (Figures XXX)

<r̸>　　LATIN SMALL LETTER RUM is used for *-rum* in *martir̸ martirum* 'martyr', *integr̸ integrum* 'intact, whole, undivided' (Figures XXX)

<ꝝ ꝝ>　　LATIN SMALL LETTER RUM ROTUNDA is used for *-rum* and *-rom* in Latin *nōꝝoꝝ nostrorum* 'of our', Portuguese *foꝝ forom* 'they went, they were' (Figures 33, 53, 61, 64, 66)

<ʀ>　　LATIN LETTER SMALL CAPITAL RUM is used for *-rum* and *-rom* in Latin *quoʀ quorum* (Figures 59, 63)

<t̸>　　LATIN SMALL LETTER TUM is used for *-tum* in *tant̸ tantum* 'so much', *quant̸ quantum* 'how much?' (Figures 58)

<ꝫ>　　LATIN SMALL LETTER UM is used for *um* and *us* in Latin '*ductibꝫ aquarꝫ*' '*ductibus aquarum*' 'to the water streams', for *os* in Latin-Portuguese *cubꝫ cubus*, *cubos* 'cubic measuring container', *neptꝫ neptos* 'grandson', and for *un* in Latin *volꝫtas voluntas* 'will', *mꝫdum mundum* 'world' (Figures XXX)

Note that the REVERSED C is not the turned c U+0254 SMALL LETTER OPEN O used in the IPA and African orthographies; it is ultimately derived from one of the Claudian letters (see §11 below). It should also be noted that these letters were widely used over a long period throughout Europe. As far west as Ireland, these conventional letters were used, sometimes for purposes quite different from their original use. The phrase *nó ro-fetatar connachta* 'or the Connachtmen found out' could be written *ꞇ rofꝶatꝗ ᷓfa*, where *ꞇ* Latin *uel* 'or' is used for Irish *nó* 'or', where the Tironian sign ⁊ is used for *et*, where *ꝗ* is used for *ar* 'on', where *ᷓ* is used for *conn* (= *coñ*), and where *r̃* Latin *sed* 'but' is used for Irish *acht* 'but'. Old Icelandic manuscripts were among the most abbreviated of all vernacular European manuscripts; in some cases almost every word in a line was abbreviated (Figure 36)

8. Combining characters. Thirteen combining superscript letters are already encoded to represent medieval Germanic manuscripts. These comprise half of the basic Latin alphabet, shown in bold type here: **a**b**cde**f**ghi**j*kl***m**n*o***pq**r*s***tuv**w**x**yz. We propose to add ten more basic superscript letters attested in medieval manuscripts which will bring the repertoire to 23 of the 26 letters: **abcdefghi**j**klmnop**q**rstuv**w**xyz**. (It should be noted that of the traditional Latin alphabet, if these ten are added, only *COMBINING LATIN SMALL LETTER J, *COMBINING LATIN SMALL LETTER Q, and *COMBINING

6

LATIN SMALL LETTER W will remain unencoded.) We also propose to encode superscripted æ, ꜵ, ꬱ, B, ç, D, ð, G, K, L, M, N, R, 2, ſ, and T.

It should be noted explicitly that the combining "capitals" in Old Norse are considered as combining *small capitals*. Thus a COMBINING SMALL CAPITAL G would be an abbreviation for <gg>, in the same manner as a LATIN LETTER SMALL CAPITAL G (on the base line, that is) would be understood as equivalent to <gg>. The relative x-height of the COMBINING SMALL CAPITALS is the same as that of the x-hight COMBINING SMALL LETTERS. The reason Old Norse added a few small capitals as superscript characters—in addition to the inventory of ordinary small characters—is the peculiar Old Icelandic custom of using small capitals for geminates; this practice was transferred to the practice of abbreviation by way of superscript characters.

<◌ᷔ>   COMBINING LATIN SMALL LETTER AE is used for *atq̊ðamikill atqvæðamikill* 'resolute' (Figures XXX)

<◌ᷕ>   COMBINING LATIN SMALL LETTER AO is used for *heĩq̊mo heimqvaomo* 'return home' (Figures XXX)

<◌ᷚ>   COMBINING LATIN SMALL LETTER AV is used for *b̊þſ føzla, brauðsføzla* 'feeding with bread' (Figures XXX)

<◌ᷗ>   COMBINING LATIN SMALL LETTER B is used for *XXXX* (Figures XXX)

<◌ᷛ>   COMBINING LATIN LETTER SMALL CAPITAL B is used for *XXXX* (Figures XXX)

<◌ᷜ>   COMBINING LATIN SMALL LETTER C CEDILLA is used for Portuguese *c̊ᵒ conçelho* 'municipality' (Figures XXX)

<◌ᷝ>   COMBINING LATIN LETTER SMALL CAPITAL D is used for *XXXX* (Figures XXX)

<◌ᷞ>   COMBINING LATIN SMALL LETTER ETH is used for *ſpɪoᷔ spiotið* 'the spear' (Figures XXX)

<◌ᷟ>   COMBINING LATIN SMALL LETTER F is used for *XXXX* (Figures XXX)

<◌ᷠ>   COMBINING LATIN SMALL LETTER G is used for *XXXX* (Figures XXX)

<◌ᷡ>   COMBINING LATIN LETTER SMALL CAPITAL G is used for *xx̊ᵒ tottogo* 'thirtieth' (Figures XXX)

<◌ᷢ>   COMBINING LATIN SMALL LETTER K is used for *m̊ mik* 'me' (Figures XXX)

<◌ᷣ>   COMBINING LATIN LETTER SMALL CAPITAL K is used for *XXXX* (Figures XXX)

<◌ᷤ>   COMBINING LATIN SMALL LETTER L is used for Latin *n̊ nihil* 'nothing', Portuguese *g̊ geral* 'general'. Old Norse *t̊ til* 'to' (Figures XXX)

<◌ᷥ>   COMBINING LATIN LETTER SMALL CAPITAL L is used for *mik̊ mikill* 'great, tall' (Figures XXX)

<◌ᷦ>   COMBINING LATIN LETTER SMALL CAPITAL M is used for *XXXX* (Figures XXX)

<◌ᷧ>   COMBINING LATIN SMALL LETTER N is used for in Latin *ů unde* 'from', *å ante* 'before', *qů quando* 'when', Old Norse *siðⁿ siðan* 'since' (Figures XXX)

<◌ᷨ>   COMBINING LATIN LETTER SMALL CAPITAL N is used for *m̊ menn* 'men' (Figures XXX)

<◌ᷩ>   COMBINING LATIN SMALL LETTER P is used for *XXXX* (Figures XXX)

<◌ᷪ>   COMBINING LATIN LETTER SMALL CAPITAL R is used for *Gunn̊ Gunnarr* 'Gunnar' (Figures XXX)

<◌ᷫ>   COMBINING LATIN SMALL LETTER R ROTUNDA is used for Latin *ᴜᴜȯ quatuor* 'four', Portuguese *p̊to porto* 'harbour', *M̊ Martim* 'Martin', Old Norse *sp̊ði spurði* 'asked' (Figures XXX)

<◌ᷲ>   COMBINING LATIN SMALL LETTER S is used for Old Norse *þ̊ þess* 'this', *h̊ hans* 'his' (Figures XXX)

<◌ᷬ>   COMBINING LATIN SMALL LETTER LONG S is used for Latin *ᴜȯ duos* 'two', *ᴜᴜé tres* 'three' (Figures XXX)

<◌ᷮ>   COMBINING LATIN LETTER SMALL CAPITAL T is used for *XXXX* (Figures XXX)

<◌ᷨ>   COMBINING LATIN SMALL LETTER Y is used for *f̊r fyrr* '*before*' (Figures XXX)

<◌ᷦ>   COMBINING LATIN SMALL LETTER Z is used for *q̊ quaðz* 'said' (Figures XXX)

In addition to these, seven other combining marks are proposed here.

<◌͆>   COMBINING DOUBLE CIRCUMFLEX ABOVE is used to denote the two diphthongs [ɛa] and [ɔu] in the first Faroese orthography by Jens Christian Svabo (1746–1824)—it is also used in editions of

7

Old English poetry to indicate disyllabic pronunciation of a diphthong that is normally monosyllabic (Figures 27, 31)

‹◌̨̇› COMBINING OGONEK ABOVE is used for marking vowel-length in Norse or to indicate vowel affection— so *ǫ̇* represents i-mutated *ø* (Figures 19, 23, 25, 28, 42)

‹◌̰› COMBINING ZIGZAG BELOW is used for *p̰, þær* 'they *f*.' together with COMBINING ZIGZAG ABOVE (Figures XXX)

‹◌̗› COMBINING IS BELOW is used in Visigothic script for *is* in *noḃ nobis* 'to us', *scripṭ scriptis* 'written', *dicṭ dictis* 'said' (Figures XXX)

‹◌̊̃› COMBINING UR ABOVE is used for *ur* in *dicir̃ dicitur* 'is said', *uocar̃ uocatur* 'is called' (Figures 32, 60, 61)

‹◌̥̊› COMBINING US ABOVE is used for medial and final *us* in *man̥ manus* 'hand', medial *os* in *p̥t post* 'after', *ꝑ̥ppitus praepositus* 'prelate, leader, governor, prevost' (Figures 32, 33, 39, 51, 52)

‹◌̆› COMBINING LATIN SMALL LETTER FLATTENED OPEN A ABOVE is used for *ua* in *q̆ qua* 'as', *ğrda guarda* 'guard', for *ra* or *ar* in Latin *con̆t contra* 'against', *sup̆ supra* 'above', Portuguese *comp̆ compra* 'a purchase', *m̆ia maria* 'Maria', *p̆te parte* 'part', for numerals *v̆ quinta* 'fifth', *ĭ prima* 'first', *una* 'one' (Figures 33, 47, 67, 73, 76).

9. Punctuation. Although many editors substitute modern for medieval punctuation, a growing number of medievalists insist that modern punctuation cannot adequately represent the syntactical features of medieval texts. In any case, at present, the reproduction of the punctuation of most medieval manuscripts using currently defined UCS characters is impossible, because many marks (e.g. PUNCTUS ELEVATUS) are simply missing. More subtly, however, even those medieval marks that have modern reflexes may differ in usage as well as in shape. For example, the modern question mark (U+003F) is descended from the PUNCTUS INTERROGATIVUS, but unlike the modern question mark, the PUNCTUS INTERROGATIVUS may be contrasted with the PUNCTUS PERCONTATIVUS—the former marking questions that require an answer while the latter marks rhetorical questions. These medieval punctuation marks are proposed below. Note that where a number of characters are described as "variants of POSITURA", this does not mean that they are freely-exchangeable glyph variants. Like the Norse ligated letters discussed in §2 above, this collection of punctuation marks is a superset of the letters found in the medieval corpus; no single manuscript contains all of them.

‹⸴› MEDIEVAL COMMA is used to mark a pause at the end of the rhetorical unit called a *comma*.

‹Γ› PARAGRAPHUS is used to mark the beginning of a section of text (Figures 53, 59).

‹⸵› POSITURA is used to mark the end of a section of text.

‹⦂⸶› COLON WITH COMMA ELEVATA POSITURA, an alternative form of POSITURA, was "employed by insular scribes at the end of a paragraph in a series of paragraphs or texts (e.g. annals) to imply that some continuation was to be expected to complete this series" (Parkes 306).

‹:⸴› COLON WITH MIDDLE COMMA POSITURA is a variant form of POSITURA.

‹⸴;› TWO DOTS OVER COMMA POSITURA is a variant form of POSITURA.

‹⸵› THREE DOTS WITH COMMA POSITURA is a variant form of POSITURA.

‹⸵› PUNCTUS ELEVATUS is used in many medieval texts to signal a wide variety of sentence-medial pauses (Figures 53, 63, 70).

‹⸮› PUNCTUS EXCLAMATIVUS (or *punctus admirativus*) is used in late medieval texts to mark the end of an exclamation.

‹⸴› PUNCTUS FLEXUS is used to mark sentence-medial pauses, but especially in liturgical texts where the pitch of the voice drops.

‹⸮› PUNCTUS INTERROGATIVUS is used to mark the end of a question that requires an answer.

‹ʔ› PUNCTUS PERCONTATIVUS is used in late texts to mark the end of a question that does not require an answer—a rhetorical question.

‹⁏› PUNCTUS VERSUS is used to mark the end of a sentence or period, or the end of a verse of a psalm.

‹.⸴› LOW PUNCTUS VERSUS is a variant form of PUNCTUS VERSUS.

<∴>   SIGNE DE RENVOI is used to connect a marginal note to a place in the text. Its size is typically largert than that of other

<:::>   SQUARED FOUR DOT PUNCTUATION.

<∴>   TWO DOTS OVER ONE DOT PUNCTUATION. (Figure 62).

<∴>   ONE DOT OVER TWO DOTS PUNCTUATION.

</>   VIRGULA SUSPENSIVA is "used to mark the briefest pause or hesitation in a text" (Parkes 307).

Three of the characters listed above are similar to three characters used in mathematics. The reason that these are proposed is to finally complete the set of generic punctuation encoded in the UCS. Most of these are based around the shape and size of the generic COLON and SEMICOLON: The UCS currently contains nine of such punctuation characters < . , ; : :· :: ∴ :: ⁙ > to which we propose to add the thirteen here: < ⁑ :∾ ;̓ ;̣. ⁑ ⁑ ⁏ , :: ⁖ ∴ >. There are also tall characters like the existing seven < ! ? / ⁝ ⁞ ⁝ ⁞ > to which we propose to add five < ⸍ ⸮ ʔ ⸪ / >. The mathematical characters U+2234 ∴ THEREFORE, U+2235 ∵ BECAUSE, U+2235 ∶ RATIO, and U+2235 ∷ PROPORTION are likely not to be based on the typical proportions of the modern COLON (indeed RATIO is *disunified* from COLON); further research needs to be done into their linebreaking and other properties.

10. Claudian letters. The Roman emperor Claudius introduced three letters to the alphabet to indicate sounds he felt could not be represented otherwise. These letters originally occurred only in capital form, but the lower-case form has been adduced to meet Unicode casing requirements as well as those of normal scholarship. An inscription may read *OLⅎMPICUS*, but the scholar may prefer to transcribe this as *Ol⊂y⊃mpicus* (as in Figure XXX) without the letter, or as *Olⅎmpicus* with it.

<Ɔ ɔ>   REVERSED C represented the consonant groups ⊂bs⊃ and ⊂ps⊃. This is the same character as the letter used for *con*, *com*, *us*, and *os* discussed in §7 below.

<ꟻ ⅎ>   INVERTED F represented consonantal ⊂v⊃ (as opposed to *u*)

<Ⱶ ⱶ>   HALF H represented the Greek ⊂y⊃ (Figures XXX).

11. Editorial punctuation. The Nordic editorial tradition uses quilled bars to mark ⱶdeletionsⱶ in the manuscript and (in reverse order) to mark ⱶdittographiesⱶ (repetitions) (Figures 26, 38, 68). The venerable *Corpus Inscriptionum Latinarum* of Berlin makes use of a large number of paired punctuation, most of which are encoded in the General Punctuation or one of the Mathematics blocks. Two pairs are not yet encoded: the SIDEWAYS U BRACKETS are used to indicate Claudian letters in a transcription, to notate inverted letters, or to notate the insertion of an image into the line of text (such as ⊂crux⊃ to indicate a cross). The ⦅double parentheses⦆ are used by some other Latinists for similar purposes. They differ from ((doubled parentheses)) quite obviously, and are not joined at the top like the WHITE PARENTHESES U+2985-U+2986 are. The FULLWIDTH WHITE PARENTHESES U+FF5F-U+FF60 used in Asia are unusual in having either a double or a joined presentation, but that glyph variation is not used by Latinists (Figures XXX).

Bibliography.

Aðalheiður Guðmundsdóttir, ed., 2001. *Úlfhams saga*. (Rit 53) Reykjavík: Stofnun Árna Magnússonar á Íslandi.

Anscombe, A., ed. 1907. "Indexes to Old-Welsh Genealogies (Continuation)", In *Archiv für Celtische Lexikographie*, vol III, ed W. Stokes & K. Meyer. [s.l.]: Max Niemeyer.

Bartholomae, Christian. 1961. *Altiranisches Wörterbuch*. 2. unveränderte Auflage. Berlin: Walter de Gruyter.

Bartoli, Francesco de. 1470. *Historia quomodo beatus Franciscus petivit a Christo indulgentiam pro Ecclesia S. Mariae de Angelis*. Trevi: Johannes Rothmann.

Brøndum-Nielsen, Johannes. [1943]. *Palæografi. A: Danmark og Sverige*. (Nordisk Kultur XXVIII:A) Stockholm: Albert Bonniers; Oslo: H. Aschehoug; København: J. H. Schultz.

Brown, Michelle P. 1993. *A Guide to Western Historical Scripts from Antiquity to 1600*. London: The British Library.

Cappelli, Adriano. 1973. *Lexicon abbreviaturarum: Dizionario di abbreviature latine ed italiane usate nelle carte e codici specialmente del medio-evo riprodotte con oltre 140000 segni incisi*. Milano: Editore Ulrico Hoepli.

Carter, Henry H., 1941. *Cancioneiro da Ajuda. A Diplomatic Edition*, New York: Modern Language Association of America / London: Oxford University Press.

Cencetti, Giorgio. 1997. *Lineamenti di storia della scrittura latina*. Seconda edizione. Pàtron Editore Bologna. ISBN 88-555-2405-4

Cleonardo, Nicolao. 1589. *Tabula in grammaticen Hebraeam auctore Nicolao Clenardo*. Lugduni Batavorum: Ex Officina Plantiniana, Apud Franciscum Raphelengium.

Dahlerup, Verner, ed., 1800. *Ágrip af Noregs konunga sögum*. Samfund til udgivelse af gammel nordisk litteratur, 2. København: Møller.

Degnbol, Helle, Bent Chr. Jacobsen, Eva Rode, Christopher Sanders, Þorbjörg Helgadóttir, eds. 1995. *Ordbog over det norrøne prosasprog. 1: a–bam*. København: Den arnamagnæanske kommission.

Eiríkur Þormóðsson & Guðrún Ása Grímsdóttir, eds. 2003. *Oddaannálar og Oddverjaannáll*. (Rit 59) Reykjavík: Stofnun Árna Magnússonar á Íslandi.

Evans, D. Simon. 1976. *A grammar of Middle Welsh*. (Medieval and Modern Welsh Series; supplementary volume) Dublin: Dublin Institute for Advanced Studies.

Farley, A. (Ed.). 1783. *Domesday Book: seu liber censualis Wilhelmi primi Regis Angliæ, inter archivos regni in domo capitulari Westmonasterii asservatus: jubente rege … Georgio Tertio prælo mandatus typis*. [London].

Firchow, Evelyn Scherabon, & Kaaren Grimstad, eds. 1989. *Elucidarius in Old Norse translation*. (Rit 36) Reykjavík: Stofnun Árna Magnússonar.

Gjerløw, Lilli, 1961: *Adoratio crucis. The Regularis Concordia and the Decreta Lanfranci. Manuscript studies in the early medieval church of Norway*. [s.l.]: Norwegian Universities Press.

Guðvarður Már Gunnlaugsson, ed. 2001. *Konungsbók Eddukvæða. Codex Regius. Stofnun Árna Magnússonar á Íslandi. Gl. Kgl. Sml. 2365 4to*. (Íslensk miðaldahandrit, 3) Reykjavík: Lögberg. ISBN: 997932161x

Haugen, Odd Einar. 1992. *Stamtre og tekstlandskap. Studiar i resensjonsmetodikk med grunnlag i Niðrstigningar saga*. 2 vols. Dr.philos. dissertation. Department of Scandinavian languages and literature, University of Bergen.

Haugen, Odd Einar, ed. 2004. *Handbok i norrøn filologi*. Bergen: Fagbokforlaget. ISBN 82-450-0105-8

Hødnebø, Finn, ed. 1960. *Corpus codicum Norvegicorum medii aevi*. Folio serie vol. II: Norske diplomer til og med år 1300. Oslo: Selskapet til utgivelse av gamle norske håndskrifter.

Holm-Olsen, Ludvig, ed., 1945. *Konungs skuggsjá*. Gammelnorske tekster, 1. Oslo: Norsk Historisk Kjeldeskrift-Institutt.

Hreinn Benediktsson, ed. 1972. *The First Grammatical Treatise: introduction, text, notes, translation, vocabulary, facsimiles*. (University of Iceland Publications in Linguistics; 1) Reykjavík: Institute of Nordic Linguistics.

Humphreys Henry Noel. [1868]. "The origin and progress of the art of writing; a connected narrative on the development of the art, in its primeval phases in Egypt, China, and Mexico; its middle state in the Cuneatic systems of Ninevah and Persepolis; its introduction to Europe through the medium of the Hebrew, Phœnician, and Greek systems; and its subsequent progress to the present day", in *Webster's Improved Dictionary of the English language, exhibiting the origin, orthography, pronunciation, & definition of words; embracing all the principal terms used in literature, science & art, according to the best authorities; and likewise giving the synonymous terms for nearly all the words explained*. 2 vols. London, Glasgow, & Edinburgh: William MacKenzie.

Jackson, Kenneth. 19xx.

John, of Garland. [ca. 1505]. *Synonima magistri* / Johannis de garlandia cu expositione magistri Galfridi anglici. [London: Wynkyn de Worde].

Johnson, Samuel. 1828. *A dictionary of the English language in which the words are deduced from their originals, and illustrated in their different significations by examples from the best writers, to which are prefixed a history of the language and an English grammar*. London: Joseph Ogle Robinson.

Jones, J. Morris. 1913. *A Welsh grammar: historical and comparative*. London: Oxford University Press.

Jones, Thomas, ed. 1941. *Brut y Tywysogyon: Peniarth Ms. 20*. Caerdydd: Gwasg Prifysgol Cymru.

[Kålund, Kristian]. 1889. *Katalog over den Arnamagnæanske Handskriftsamling. Udgivet af Kommissionen for det Arnamagnæanske Legat*. København: Gyldendalske Boghandel.

Klaeber, F., ed., 1950. *Beowulf and the Fight at Finnsburg*. 3rd ed. Boston: D.C. Heath.

Konráð Gíslason. 1846. *Um frum-parta íslenzkrar túngu í fornöld*. Kaupmannahöfn.

Maia, Clarinda de Azevedo. 1986. *História do galego-português. Estado linguístico da Galiza e do Noroeste de Portugal do século XIII ao século XVI*. Coimbra: Instituto Nacional de Investigação Científica.

Marín Martínez, Tomás. 1991. *Paleografíca y diplomática*. Quinta edición. 1999 sexta reimpresión. Universidad Nacional de Educación a Distancia (UNED). ISBN 84-362-2052-8

Matras, Christian, ed. 1939. *Svabos færøske Visehaandskrifter.* (Samfund til udgivelse af gammel nordisk litteratur, 59) København: Bianco Lunos Bogtrykkeri A/S.

Millares Carlo, Agustín. 1983. *Tratado de Paleografía Española*. Madrid: Espasa-Calpe

Nunes, Eduardo Borges. 1969. *Álbum de Paleografia Portuguesa*. Vol. I. Lisboa: Instituto de Alta Cultura / Centro de Estudos Históricos.

Ó Cuiv, Brian, ed. 1994. *Aibidil Gaoidheilge & Caiticiosma: Seaán Ó Cearnaigh's Irish Primer of Religion published in 1571*. Dublin: Dublin Institute for Advanced Studies. ISBN 1-85500-163-2

Pacheco, José. 1988. *A Divina Arte Negra e o Livro Português (séculos XV e XVI)*, Lisboa: Vega, Fig. 3 — Folha do Prólogo do Sacramental de Clemente Sanchez, Chaves, Autor desconhecido, [1488]., Aquele que se pressupõe ter sido, em Portugal, o primeiro livro impresso em português., p. 87

Parkes, M. B. 1992. *Pause and Effect: An Introduction to the History of Punctuation in the West*. Berkeley: University of California Press. ISBN 0-520-07941-8

Santos, Maria José de Azevedo. 1994. *Da Visigótica à Carolina: a escrita em Portugal de 882 a 1172*. Lisboa: Fundação Calouste Gulbenkian, Junta Nacional de Investigação Científica e Tecnológica. ISBN 84-376-1245-4

Saxoferrato, Bartholus de. 1471. *Lectura super I. parte Infortiati*. Trevi: Johannes Rothmann.

Stefán Karlsson, ed. 1963. *Islandske originaldiplomer indtil 1450. Tekst*. (Editiones Arnamagnæanæ, Series A, vol. 7) København: Munksgaard.

Tertullian, Quintus Septimus Florens. [1493]. *Apologeticus adversus gentes*. Venetiis: B. Benalius.

Thomas, Graham, and Nicholas Williams, eds. [In preparation]. *Bewnans Ke: The Life of St Kea*. Aberystwyth: National Library of Wales.

van Arkel-de Leeuw van Weenen, Andrea, ed. 1987. *Möðruvallabók: AM 132 Fol*. Volume Two: Text. Leiden: E. J. Brill.

Virgile. 1509. *Opera com. de Servius*. Milano: Leonardo Pachel

Acknowledgements

	xDC	xDD	xDE	xDF
0		◌ꝋ xxD0	◌L xxE0	
1		◌̃ xxD1	◌M xxE1	
2		◌æ xxD2	◌n xxE2	
3		◌ao xxD3	◌N xxE3	
4		◌aͬ xxD4	◌p xxE4	
5		◌b xxD5	◌R xxE5	
6		◌B xxD6	◌2 xxE6	
7		◌ç xxD7	◌s xxE7	
8		◌D xxD8	◌f xxE8	
9		◌ð xxD9	◌T xxE9	
A		◌f xxDA	◌y xxEA	
B	◌͡ xxCB	◌g xxDB	◌z xxEB	
C	◌ꝛ xxCC	◌G xxDC		
D	◌ xxCD	◌k xxDD		
E	◌ xxCE	◌K xxDE		
F	◌ xxCF	◌l xxDF		

Note to reviewers: This is a preliminary proposal. The code points shown are merely suggestions for a possible placement.

Medievalist and Latinist additions

1DCB ◌͛ COMBINING DOUBLE CIRCUMFLEX ABOVE
1DCC ◌ COMBINING OGONEK ABOVE
1DCD ◌ COMBINING ZIGZAG BELOW
1DCE ◌ COMBINING IS BELOW
1DCF ◌ COMBINING UR ABOVE
1DD0 ◌ COMBINING US ABOVE
→ A764 ꝥ modifier letter us

Superscript Latin letter diacritics

1DD1 ◌ COMBINING LATIN SMALL LETTER FLATTENED OPEN A
1DD2 ◌ COMBINING LATIN SMALL LETTER AE
1DD3 ◌ COMBINING LATIN SMALL LETTER AO
1DD4 ◌ COMBINING LATIN SMALL LETTER AV
1DD5 ◌ COMBINING LATIN SMALL LETTER B
1DD6 ◌ COMBINING LATIN LETTER SMALL CAPITAL B
1DD7 ◌ COMBINING LATIN SMALL LETTER C CEDILLA
1DD8 ◌ COMBINING LATIN LETTER SMALL CAPITAL D
1DD9 ◌ COMBINING LATIN SMALL LETTER ETH
1DDA ◌ COMBINING LATIN SMALL LETTER F
1DDB ◌ COMBINING LATIN SMALL LETTER G
1DDC ◌ COMBINING LATIN LETTER SMALL CAPITAL G
1DDD ◌ COMBINING LATIN SMALL LETTER K
1DDE ◌ COMBINING LATIN LETTER SMALL CAPITAL K
1DDF ◌ COMBINING LATIN SMALL LETTER L
1DE0 ◌ COMBINING LATIN LETTER SMALL CAPITAL L
1DE1 ◌ COMBINING LATIN LETTER SMALL CAPITAL M
1DE2 ◌ COMBINING LATIN SMALL LETTER N
1DE3 ◌ COMBINING LATIN LETTER SMALL CAPITAL N
1DE4 ◌ COMBINING LATIN SMALL LETTER P
1DE5 ◌ COMBINING LATIN LETTER SMALL CAPITAL R
1DE6 ◌ COMBINING LATIN SMALL LETTER R ROTUNDA
1DE7 ◌ COMBINING LATIN SMALL LETTER S
1DE8 ◌ COMBINING LATIN SMALL LETTER LONG S
→ 017F ſ latin small letter long s
1DE9 ◌ COMBINING LATIN LETTER SMALL CAPITAL T
1DEA ◌ COMBINING LATIN SMALL LETTER Y
1DEB ◌ OMBINING LATIN SMALL LETTER Z

	xE8	xE9	xEA	xEB	xEC	xED	xEE	xEF
0								
1								
2								
3								
4								
5								
6								
7								
8								
9								
A								IL xxFA
B								tt xxFB
C		f xx9C						ʋ xxFC
D		f xx9D						ɓ xxFD
E								y xxFE
F								y xxFF

Note to reviewers: This is a preliminary proposal. The code points shown are merely suggestions for a possible placement.

Latinist and medievalist characters

1E9C ꞔ LATIN SMALL LETTER LONG S WITH DIAGONAL
STROKE
→ 017F ſ latin small letter long s

1E9D ẝ LATIN SMALL LETTER LONG S WITH HIGH
STROKE

Latinist and medievalist characters

1EFA Ỻ LATIN CAPITAL LETTER MIDDLE-WELSH LL

1EFB ỻ LATIN SMALL LETTER MIDDLE-WELSH LL

1EFC Ỽ LATIN CAPITAL LETTER MIDDLE-WELSH V
→ 01B2 Ʋ latin capital letter v with hook

1EFD ỽ LATIN SMALL LETTER MIDDLE-WELSH V
→ 028B ʋ latin small letter v with hook

1EFE Ỿ LATIN CAPITAL LETTER Y WITH LOOP

1EFF ỿ LATIN SMALL LETTER Y WITH LOOP

	xC6	xC7
0		
1		
2		
3		
4		
5		
6		
7		
8		
9		
A		
B	δ xx6B	
C		
D		
E		
F		

Note to reviewers: This is a preliminary proposal. The code points shown are merely suggestions for a possible placement.

Latinist and medievalist character

2C6B   δ   LATIN SMALL LETTER SCRIPT D
       → 018D ϙ   latin small letter turned delta
       → 03B4 δ   greek small letter delta

	xE0	xE1	xE2	xE3	xE4	xE5	xE6	xE7
0			⊦ xx20	ʔ xx30	∴ xx40			
1			⊣ xx21	Γ xx31	∴ xx41			
2			⊂ xx22	ʔ xx32	/ xx42			
3			⊃ xx23	∴ xx33				
4			⦅ xx24	∴ xx34				
5			⦆ xx25	∴ xx35				
6			Γ xx26	∴ xx36				
7			⌐ xx27	∴ xx37				
8			L xx28	∮ xx38				
9			⌐ xx29	∴ xx39				
A				ʔ xx3A				
B				ʕ xx3B				
C				ʔ xx3C				
D				ʒ xx3D				
E		○ xx1E		∴ xx3E				
F		⫰ xx1F		∴ xx3F				

Note to reviewers: This is a preliminary proposal. The code points shown are merely suggestions for a possible placement.

Latinist and medievalist editorial punctuation

2E1E ˚ MIDDLE RING
　→ 00B0 ° degree sign
　→ 02DA ˚ ring above
　→ 1424 ᐤ candian syllabics final ring
　→ 2218 ∘ ring operator
　→ 25E6 ◦ white bullet
　→ 26AC ○ medium small white circle

2E1F ⸟ PALM BRANCH
　= ramus

2E20 ⸠ LEFT VERTICAL BAR WITH QUILL

2E21 ⸡ RIGHT VERTICAL BAR WITH QUILL

2E22 ⸢ LEFT SIDEWAYS U BRACKET
　→ 2282 ⊂ subset of

2E23 ⸣ RIGHT SIDEWAYS U BRACKET
　→ 2283 ⊃ superset of

2E24 ⸤ LEFT DOUBLE PARENTHESIS
　→ 2985 ⦅ left white parenthesis
　→ FF5F ⦅ fullwidth left white parenthesis

2E25 ⸥ RIGHT DOUBLE PARENTHESIS
　→ 2986 ⦆ left right parenthesis
　→ FF60 ⦆ fullwidth right white parenthesis

2E26 ⸦ TOP LEFT HALF SQUARE BRACKET
　→ 2308 ⌈ left ceiling
　→ 231C ⌜ top left corner
　→ 2E00 ⸀ right angle substitution marker
　→ 300C 「 left corner bracket

2E27 ⸧ TOP RIGHT HALF SQUARE BRACKET
　→ 2309 ⌉ right ceiling
　→ 231D ⌝ top right corner

2E28 ⸨ BOTTOM LEFT HALF SQUARE BRACKET
　→ 230A ⌊ left floor
　→ 231E ⌞ bottom left corner

2E29 ⸩ BOTTOM RIGHT HALF SQUARE BRACKET
　→ 230B ⌋ right floor
　→ 231F ⌟ bottom right corner
　→ 300D 」 right corner bracket

Medieval punctuation

2E30 ⸰ MEDIEVAL COMMA

2E31 ⸱ PARAGRAPHUS
　→ 00B6 ¶ pilcrow sign
　→ 204B ⁋ reversed pilcrow sign
　→ 2761 ❡ curved stem paragraph sign ornament
　→ 2E0F ⸏ paragraphos

2E32 ⸲ POSITURA

2E33 ⸳ COLON WITH COMMA ELEVATA POSITURA

2E34 ⸴ COLON WITH MIDDLE COMMA POSITURA

2E35 ⸵ TWO DOTS OVER COMMA POSITURA

2E36 ⸶ THREE DOTS WITH COMMA POSITURA

2E37 ⸷ PUNCTUS ELEVATUS

2E38 ⸸ PUNCTUS EXCLAMATIVUS

2E39 ⸹ PUNCTUS FLEXUS

2E3A ⸺ PUNCTUS INTERROGATIVUS

2E3B ⸻ PUNCTUS PERCONTATIVUS
　→ 2426 ␦ symbol for substitute form two
　→ 061F ؟ arabic question mark

2E3C ⸼ PUNCTUS VERSUS

2E3D ⸽ LOW PUNCTUS VERSUS

2E3E ⸾ SIGNE DE RENVOI
　→ 205A ⁚ two dot punctuation
　→ 205B ⁛ four dot mark
　→ 2237 ∷ proportional

2E3F ⸿ FOUR DOTS PUNCTUATION
　→ 2058 ⁘ four dot punctuation
　→ 2059 ⁙ five dot punctuation

2E40 ⹀ TWO DOTS OVER ONE DOT PUNCTUATION
　→ 10FB ჻ georgian paragraph separator
　→ 2056 ⁖ three dot punctuation
　→ 2235 ∵ because

2E41 ⹁ ONE DOT OVER TWO DOTS PUNCTUATION
　→ 2234 ∴ therefore

2E42 ⹂ VIRGULA SUSPENSIVA

	x72	x73	x74	x75	x76	x77	x78	x79	x7A	x7B	x7C	x7D	x7E	x7F
0	AA xx20	Ⱶ xx30	Oᵒ xx40	2ꝶ xx50	ſ xx60	rꝰ xx70								
1	æ xx21	Ɉ xx31	oᵒ xx41	2ꝶ xx51	ſ xx61	Rꝰ xx71								
2	AO xx22	Ⱶ xx32	OO xx42	Ỽ xx52	9 xx62	ꝥ xx72								
3	ao xx23	ƕ xx33	oo xx43	ỽ xx53	9 xx63	ꝳ xx73								
4	AꞘ xx24	Ꝁ xx34	P xx44	Ỽ xx54	9 xx64									
5	aꞙ xx25	ꝁ xx35	ꝑ xx45	ỽ xx55	ꝺ xx65									
6	AꞘ xx26	Ꝁ xx36	Ꝑ xx46	Ꝣ xx56	Ꝼ xx66									
7	aꞙ xx27	ꝁ xx37	ꝑ xx47	ꝣ xx57	ꝼ xx67									
8	AꞘ xx28	Ꝁ xx38	Ꝓ xx48	Ꝧ xx58	ꞁ xx68									
9	aꞙ xx29	ꝁ xx39	ꝓ xx49	ꝧ xx59	ꞁ xx69									
A	AꞘ xx2A	Ꝇ xx3A	Q xx4A	Ꝥ xx5A	ꞇ xx6A									
B	aꞙ xx2B	ꝇ xx3B	q xx4B	ꝥ xx5B	F xx6B									
C	Ꜿ xx2C	Ł xx3C	Ꝙ xx4C	Ỿ xx5C	s xx6C									
D	ꜿ xx2D	ꝉ xx3D	ꝙ xx4D	ỿ xx5D	ꝏ xx6D									
E	Ꜩ xx2E	Ꝋ xx3E	2 xx4E	3 xx5E	mꝰ xx6E									
F	ꜩ xx2F	ꝋ xx3F	2 xx4F	3 xx5F	mꝰ xx6F									

Note to reviewers: This is a preliminary proposal. The code points shown are merely suggestions for a possible placement.

Medievalist additions

A720   ꜡   LATIN CAPITAL LETTER AA
A721   ꜡   LATIN SMALL LETTER AA
A722   Ꜣ   LATIN CAPITAL LETTER AO
A723   ꜣ   LATIN SMALL LETTER AO
A724   Ꜥ   LATIN CAPITAL LETTER AU
A725   ꜥ   LATIN SMALL LETTER AU
A726   Ꜧ   LATIN CAPITAL LETTER AV
A727   ꜧ   LATIN SMALL LETTER AV
A728   Ꜩ   LATIN CAPITAL LETTER AV WITH HORIZONTAL BAR
A729   ꜩ   LATIN SMALL LETTER AV WITH HORIZONTAL BAR
A72A   Ꜫ   LATIN CAPITAL LETTER AY
A72B   ꜫ   LATIN SMALL LETTER AY
A72C   Ɔ   LATIN CAPITAL LETTER REVERSED C
       = capital antisigma
       → 2183 Ɔ roman numeral reversed one hundred
       → 03FD Ϲ greek capital lunate sigma symbol
A72D   ɔ   LATIN SMALL LETTER REVERSED C
       = small antisigma
A72E   Ꜿ   LATIN CAPITAL LETTER REVERSED C WITH DOT
       → 03FF Ͽ greek capital reversed dotted lunate sigma symbol
A72F   ꜿ   LATIN SMALL LETTER REVERSED C WITH DOT
A730   Ⅎ   LATIN CAPITAL LETTER TURNED F
       = capital digamma inversum
       → 2132 Ⅎ turned capital f
A731   Ꞁ   LATIN SMALL LETTER TURNED F
       = small digamma inversum
       → 025F ɟ latin small letter dotless j with stroke
A732   Ħ   LATIN CAPITAL LETTER HALF H
A733   ħ   LATIN SMALL LETTER HALF H
A734   Ꝃ   LATIN CAPITAL LETTER K WITH STROKE
A735   ꝃ   LATIN SMALL LETTER K WITH STROKE
A736   Ꝅ   LATIN CAPITAL LETTER K WITH DIAGONAL STROKE
A737   ꝅ   LATIN SMALL LETTER K WITH DIAGONAL STROKE
A738   Ꝁ   LATIN CAPITAL LETTER K WITH STROKE AND DIAGONAL STROKE
A739   ꝁ   LATIN SMALL LETTER K WITH STROKE AND DIAGONAL STROKE
A73A   Ꝉ   LATIN CAPITAL LETTER BROKEN L
A73B   ꝉ   LATIN SMALL LETTER BROKEN L
A73C   Ꝇ   LATIN CAPITAL LETTER L WITH HIGH STROKE
       → 0141 Ł latin capital letter l with stroke
A73D   ꝇ   LATIN SMALL LETTER L WITH HIGH STROKE
       → 0142 ł latin small letter l with stroke
A73E   Ꝋ   LATIN CAPITAL LETTER O WITH LONG STROKE OVERLAY
A73F   ꝋ   LATIN SMALL LETTER O WITH LONG STROKE OVERLAY
A740   Ơ   LATIN CAPITAL LETTER O WITH LOOP
       → 01A0 Ơ latin capital letter o with horn
A741   ơ   LATIN SMALL LETTER O WITH LOOP
       → 01A1 ơ latin small letter o with horn
A742   Ꝏ   LATIN CAPITAL LETTER OO
A743   ꝏ   LATIN SMALL LETTER OO
A744   Ꝑ   LATIN CAPITAL LETTER P WITH STROKE THROUGH DESCENDER

A745   ꝑ   LATIN SMALL LETTER P WITH STROKE THROUGH DESCENDER
A746   Ꝓ   LATIN CAPITAL LETTER P WITH FLOURISH
A747   ꝓ   LATIN SMALL LETTER P WITH FLOURISH
A748   Ꝕ   LATIN CAPITAL LETTER P WITH SQUIRREL TAIL
A749   ꝕ   LATIN SMALL LETTER P WITH SQUIRREL TAIL
A74A   Ꝗ   LATIN CAPITAL LETTER Q WITH STROKE THROUGH DESCENDER
A74B   ꝗ   LATIN SMALL LETTER Q WITH STROKE THROUGH DESCENDER
A74C   Ꝙ   LATIN CAPITAL LETTER Q WITH DIAGONAL STROKE
A74D   ꝙ   LATIN SMALL LETTER Q WITH DIAGONAL STROKE
A74E   Ꝛ   LATIN CAPITAL LETTER R ROTUNDA
       → 0032 2 digit two
       → 01BB ƻ latin letter two with stroke
A74F   ꝛ   LATIN SMALL LETTER R ROTUNDA
A750   Ꝝ   LATIN CAPITAL LETTER RUM ROTUNDA
A751   ꝝ   LATIN SMALL LETTER RUM ROTUNDA
A752   Ꝟ   LATIN CAPITAL LETTER V WITH DIAGONAL STROKE
       → 2123 ℣ versicle
A753   ꝟ   LATIN SMALL LETTER V WITH DIAGONAL STROKE
A754   Ꝡ   LATIN CAPITAL LETTER VY
A755   ꝡ   LATIN SMALL LETTER VY
A756   Ꝣ   LATIN CAPITAL LETTER VISIGOTHIC Z
A757   ꝣ   LATIN SMALL LETTER VISIGOTHIC Z
A758   Ꝥ   LATIN CAPITAL LETTER THORN WITH STROKE
A759   ꝥ   LATIN SMALL LETTER THORN WITH STROKE
A75A   Ꝧ   LATIN CAPITAL LETTER THORN WITH STROKE THROUGH DESCENDER
A75B   ꝧ   LATIN SMALL LETTER THORN WITH STROKE THROUGH DESCENDER
A75C   Ꝩ   LATIN CAPITAL LETTER VEND
       0x 0194 Ɣ latin capital letter gamma
A75D   ꝩ   LATIN SMALL LETTER VEND
A75E   Ꝫ   LATIN CAPITAL LETTER ET
       → 01B7 Ʒ latin capital letter ezh
A75F   ꝫ   LATIN SMALL LETTER ET
       → 0292 ʒ latin small letter ezh
A760   Ꝭ   LATIN CAPITAL LETTER IS
A761   ꝭ   LATIN SMALL LETTER IS
A762   Ꝯ   LATIN CAPITAL LETTER CON
A763   ꝯ   LATIN SMALL LETTER CON
A764   ꝰ   MODIFIER LETTER US

Insular tradition

Other characters for the insular tradition are coded elsewhere

       → 1D79 ᵹ latin small letter insular g
A765   ꝺ   LATIN SMALL LETTER INSULAR D
       → 00F0 ð latin small letter eth
A766   Ꝼ   LATIN CAPITAL LETTER INSULAR F
A767   ꝼ   LATIN SMALL LETTER INSULAR F
A768   ꝸ   LATIN SMALL LETTER INSULAR R
A769   ꝸ   LATIN SMALL LETTER INSULAR S
A76A   ꝼ   LATIN SMALL LETTER INSULAR T

Other additions

Other small capital letters for phonetic transcription are encoded elsewhere

A76B   ꜰ   LATIN LETTER SMALL CAPITAL F

A76C   Ꝭ   LATIN LETTER SMALL CAPITAL S
         → 017F ſ   latin small letter long s

Latinist abbreviations

A76D   ꝭ   LATIN SMALL LETTER DUM
A76E   Ꝯ   LATIN SMALL LETTER MUM
A76F   ꝯ   LATIN SMALL LETTER NUM
A770   ꝰ   LATIN SMALL LETTER RUM
         → 211E ℞   prescription take
         → 211F ℟   response
A771   ꝱ   LATIN LETTER SMALL CAPITAL RUM
A772   ꝲ   LATIN SMALL LETTER TUM
A773   ꝳ   LATIN SMALL LETTER UM