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 five Latin Tironian letters to the UCS**
Source:	**Michael Everson and Andrew West**
Status:	**Individual Contribution**
Date:	**2017-09-04**

0. Summary. This proposal requests the encoding of a casing partner to one existing UCS character, the change of some properties of that existing UCS character, and the addition of two new casing letters used in medieval European texts. If this proposal is accepted, the following characters will exist:

⅂ 204A TIRONIAN SIGN ET

    • Irish Gaelic, Old English, …

    • typically drawn at x-height with a slightly angled full-length descender

    → 0026 & ampersand

    → A7F0 tironian sign capital et

    → 1F670 🙰 script ligature et ornament

⅂ A7F0 LATIN CAPITAL LETTER TIRONIAN ET

    → 204A tironian sign et

ꝼ A7F1 LATIN CAPITAL LETTER TIRONIAN ET WITH HOOK

    • Middle English, …

ꝼ A7F2 LATIN SMALL LETTER TIRONIAN ET WITH HOOK

ꝼ A7F3 LATIN CAPITAL LETTER TIRONIAN ET WITH HOOK AND STROKE

    • Middle English, Latin, …

ꝼ A7F4 LATIN SMALL LETTER TIRONIAN ET WITH HOOK AND STROKE

1. Background. A variety of medieval manuscripts across a number of centuries treat the TIRONIAN SIGN ET as a letter of the alphabet, and cases it when in sentence-initial position. Modern transcribers of documents containing these letters have distinguished them as casing, sometimes using the digit 7 as a font workaround. The simple two-stroke shape of the character ⅂ as used in the insular tradition (for Irish and Old English) was lost in time to a form with a hooked base, sometimes with a long extension of the topbar encircling the glyph. Another descendent of the original insular character has a strong horizontal stroke through it; this may or may not have the long hooked stroke, or it may have a very short top stroke entirely. Modern editors distinguish the glyphs of the insular, hooked, and hooked-with-stroke form even in a single publication.

2. Property changes for U+204A. Encoded in the General Punctuation block, U+204A will require some property changes for this proposal to be successful. The current property values for U+204A are:

```
gc=Po, Other_Punctuation
ccc=0, Not_Reordered
bc=N, Other_Neutral
lb=AL (Alphabetic)
Mirrored=No
Script=Common
Grapheme_Base=Y [derived]
```

Proposed property values for U+204A are:

```
gc=Ll, Lowercase_Letter
ccc=0, Not_Reordered
bc=L, Left_To_Right
lb=AL (Alphabetic)
Mirrored=No
Script=Latin
Changes_When_Uppercased=Y [derived]
Alphabetic=Y [derived]
Changes_When_Titlecased=Y [derived]
Lowercase=Y [derived]
Grapheme_Base=Y [derived]
Changes_When_Casemapped=Y [derived]
Cased=Y [derived]
```

With regard to the Pattern_Syntax value, it is True for U+204A because it is in the General Punctuation block. This value cannot be changed due to stability policy. Because of the Pattern_Syntax value for U+204A, even after the change of General_Category to Ll, U+204A will not be usable in identifiers.

3. Usage. These characters are typically used to represent the Latin word *et*, the Old English word *ond*, and the Irish word *agus*, all of which mean 'and'. In Irish *only* the paradigmatic insular glyph ⁊ is used—by modern Irish users, not just by medievalists—and the hooked glyphs and barred glyphs of ⁊⁊ and ⁊⁊ are unknown and simply would not be recognized by anyone. Unification of the hooked characters with or without bars would result in illegibility in Irish texts (and public signage), and is therefore inappropriate. Just as significant, however, is the distinction presses like Oxford have made of the three letters, particularly where discussion the provenance and age of a manuscript in palaeographic studies is relevant. Glyphic variation in the hooked glyphs includes swash forms like ⁊ and ⁊ (both forms occurring also with the horizontal stroke), or like ⁊, but modern editors do not distinguish these from the regular hooked glyphs without swash.

4. Glyphs. A variety of glyph shapes are available for the hooked TIRONIAN ETS, as can be seen in the Figures below. For the reference glyphs, we have chosen a typographic hybrid of the bow of an inverted digit 2 and the crossbar of a digit 7, as shown in Figure 12: ⁊⁊⁊. Constructing the hooked glyphs from such pieces will usually yield an attractive glyph harmonious with the base font. It should be noted, however, that since U+204A was encoded in Unicode 3.0. many implementations of it have been seriously incorrect. The typically bad glyph is no more than an x-height down-and-left right angle, as in Lucida Grande ⁊. The correct shape has a slightly thicker horizontal at x-height, often slightly bowed downwards, and with a thinner vertical going down to the descender length of the letter *p*, usually offset towards the left from true vertical. The glyph should never be identical to digit 7: ⁊⁊.

5. Identifiers. As the characters proposed here are historical characters, we expect that none of these six characters would be required in IDNA identifiers.

6. Ordering. The Old English alphabet was formally ordered by Byrhtferð in the year 1011 for the purposes of numerology. His alphabetical order included both AMPERSAND & and the TIRONIAN SIGN ⅂.

A B C D E F G H I K L M N O P Q R S T V X Y Z & ⅂ ꝥ Þ Ð Æ

Today's default UCS order is somewhat different. The ⅂ TIRONIAN SIGN ET is ordered as punctuation, after & AMPERSAND:

… / < \ < & < ⅂ < # < % …

Since the Tironian marks are related in a certain sense to some of the Latin abbreviation letters, it seems sensible to put them, interfiled, between Norse Ⅴ VEND (which derives from Ƿ WYNN) and ⳍ LATIN LETTER ET:

A << (Æ) < B < C < D << Ð < E < F < G < H < I < J < K < L < M < N < O < P < Q < R < S < T < U < V < X < Y < Z < Þ < Ƿ < Ⅴ < ⅂ <<< ꝛ << ꝛ <<< ꝛ << ꝛ <<< ꝛ < ⳍ < ꝓ …

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

```
A7F0;TIRONIAN SIGN ET;Ll;0;L;;;;;N;;;A7F0;;A7F0
A7F0;LATIN CAPITAL LETTER TIRONIAN ET;Lu;0;L;;;;;N;;;;20A4;
A7F1;LATIN CAPITAL LETTER TIRONIAN ET WITH HOOK;Lu;0;L;;;;;N;;;;A7F2;
A7F2;LATIN SMALL LETTER TIRONIAN ET WITH HOOK;Ll;0;L;;;;;N;;;A7F1;;A7F1
A7F3;LATIN CAPITAL LETTER TIRONIAN ET WITH HOOK AND STROKE;Lu;0;L;;;;;N;;;;A7F4;
A7F4;LATIN SMALL LETTER TIRONIAN ET WITH HOOK AND STROKE;Ll;0;L;;;;;N;;;A7F3;;A7F3
```

8. Bibliography

Conner, Patrick W. 1993. *Anglo-Saxon Exeter: A Tenth-century Cultural History*. Woodbridge: Boydell Press. ISBN 978-0-85115-307-0

Parkes, M. B. 1979. *English Cursive Book Hands 1250-1500*. London: Scolar Press. ISBN 0-85967-535-1

Parkes, M. B. 1993. *Pause and effect: an introduction to the history of punctuation in the West*. Berkeley & Los Angeles: University of California Press. ISBN 0-520-07941-8.

Wright, Andrew. 1879. *Court-Hand Restored or, The Student's Assistant in reading Old Deeds, Charters, Records, etc*. Ninth edition, ed. Charles Trice Martin. London: Reeves & Turner.

Wright, C. E. 1960. *English vernacular hands from the twelfth to the fifteenth centuries*. (Oxford Palaeographical Handbooks.) Oxford: Clarendon Press.
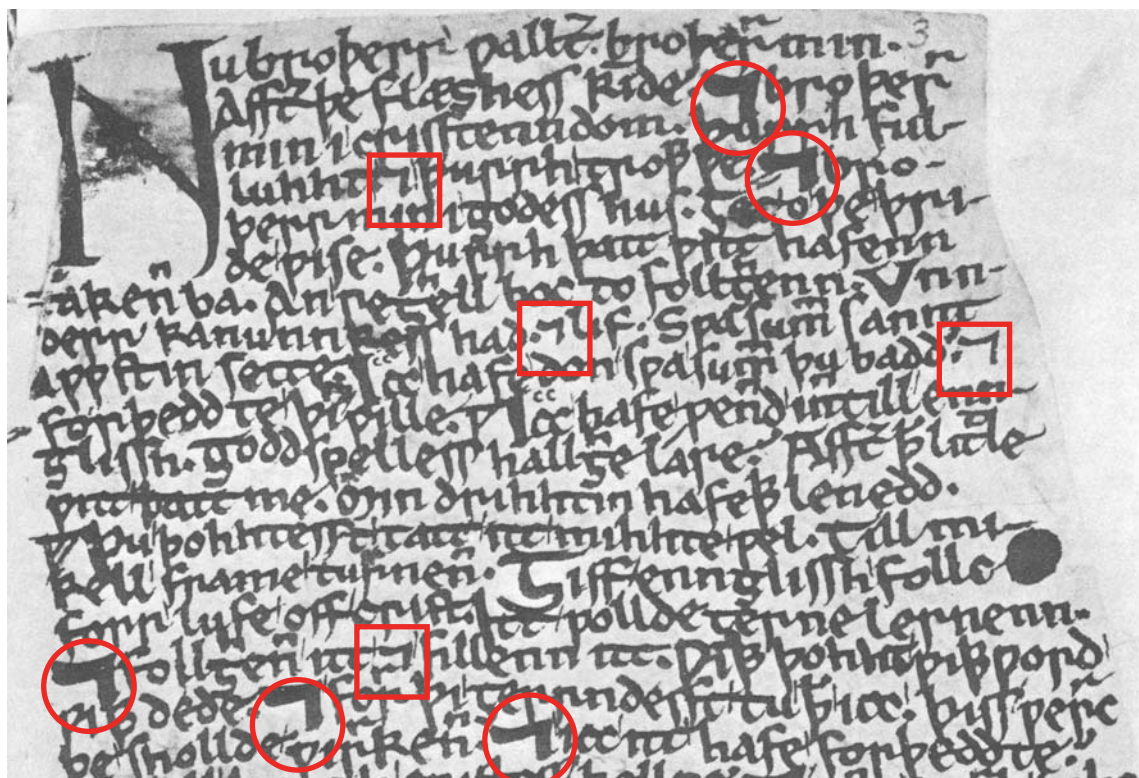
Figures.

Figure 1. Old English text, "The Ormulum', written early 13th century, Oxford Bodleian Library, Junius MS 1, fol. 3r. At the beginnings of sentences the letters A, I, V, Ƿ, ȝ, and Þ are given alongside capital ⁊. The Ormulum is written in a very precise orthography and is an extraordinarily important source of our knowledge of Middle English pronunciation. From Wright 1960, Plate 2.

Figure 2. Old English text, "The Ormulum', written early 13th century, Oxford Bodleian Library, Junius MS 1, fol. 3r. In this transcription a distinction between capital ⁊ and small ⁊ has not been made, likely due to typographic constraints. From Wright 1960, Plate 2.
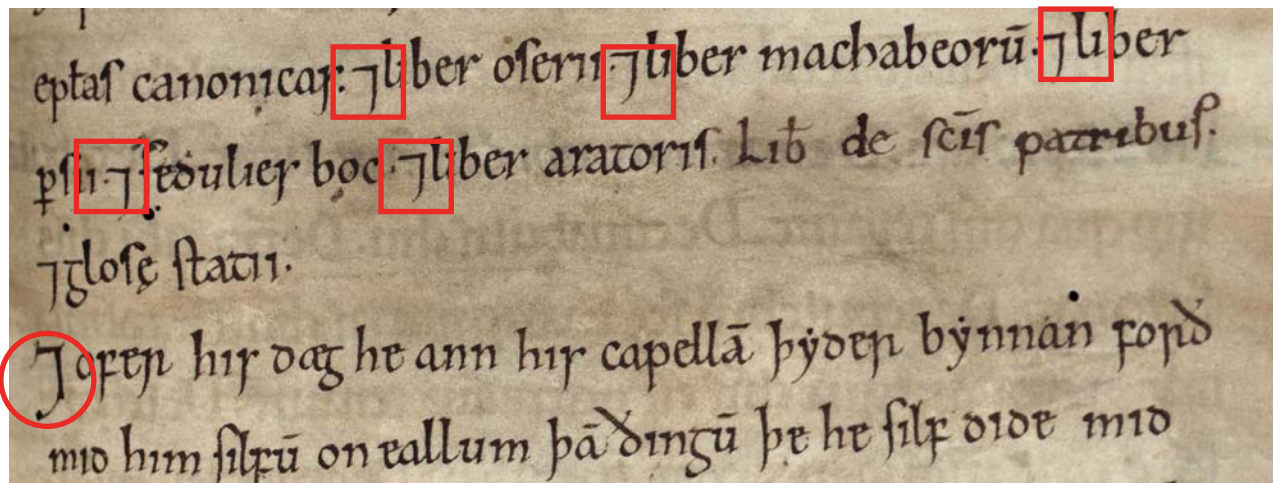
Figure 3. Latin and Old English text, "The Gifts of Bishop Leofric', written early 13th century, Exeter Cathedral Library, MS 3501, fol. 2r. At the beginnings of sentences the capital ꝛ can clearly be seen. (There are other examples of this in the manuscript as well.)

Figure 4. Transcription of the Bishop Leofric text from Conner 1993. Here, with 7-shaped glyphs lowercase ⁊ and uppercase ꝛ are distinguished.

Figure 5. Transcription of another MS containing the Bishop Leofric text, from an edition of the Anglo-Saxon charters. Here both lowercase ⁊ and uppercase ꝛ are clearly distinguished.

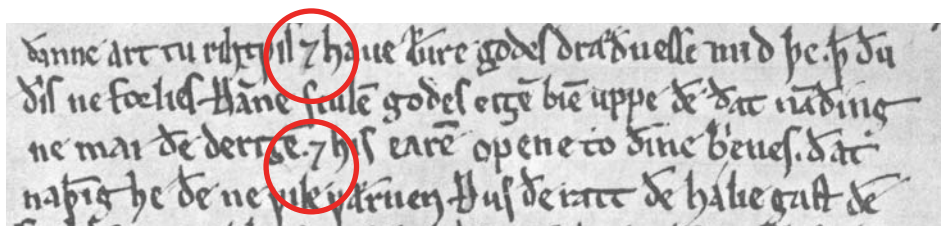**Figure 6.** Translation of the text in Figure 5.

Figure 7. Middle English text, "Vices and Virtues', written c. 1200, British Museum, Stowe MS 34, fol. 20r. Here the early hooked form of ꝛ TIRONIAN ET WITH HOOK can be seen. From Wright 1960, Plate 3.

'VICES AND VIRTUES', the earliest of all Middle English dialogues, composed in the Essex dialect, about A.D. 1200.

Written about A.D. 1200.

Vellum: 22 × 16 cm.

A vigorous, individualistic hand closely related to the court-hand of the time.

Note. The vertical stroke of the *a* is normally now brought well up but some examples preserve hint of the insular *a* (ᴧ); a remarkable form of *æ* is used, the *e* being suprascript and linked to the top part of the vertical of the *a*; long *s* appears throughout; continental *f* and *r* are now normal; after *o* the 2-form of *r* is used; 'yogh' (ȝ) still preserves in its angularity traces of its origin in the insular *g* (ᵹ); hard *g* is represented by what is now to be the normal form; the insular or uncial *d* (ð) is used; þ and ð are used; þ and p are clearly differentiated; *c* and *t* are well differentiated, the cross-bar of the latter being firmly made (and the vertical does not yet appear above it); *i* has sometimes an upward slanting stroke.

Abbreviations are rare: þ is used for *þæt* and the Tironian nota for *and* is used (⁊) (in a rather elaborate form ꝛ); the horizontal mark of contraction is sometimes hooked at the end.

Correction is by a dot under the letter to be deleted (e.g. *f* in l. 12); this is normal medieval practice.

Text printed in: *Vices and Virtues*, ed. F. Holthausen (E.E.T.S., O.S., nos. 89, 159 (1888, 1921)).

Another specimen in: Pal. Soc., ser. ii, pl. 94.

Our plate shows fol. 20ʳ.

TRANSCRIPTION

ðanne art tu rihtꝬis ꝛ haue æure godes drædnesse mid þe . þ ðu
ðis ne forlies Ðanc seȝ̄e godes eiȝē biē uppe ðe ðat nāðíng
ne mai ðe deriȝē . ꝛ his earē opene to ðine bʼenes . ðat
naþīg he ðe ne pͥle parnen Ðus ðe ratt ðe halie gast ðe

Figure 8. Transcription of Figure 7, showing ꝛ TIRONIAN ET WITH HOOK. In the discussion both the forms ⁊ and ꝛ are distinguished by the editor. From Wright 1960, Plate 3.
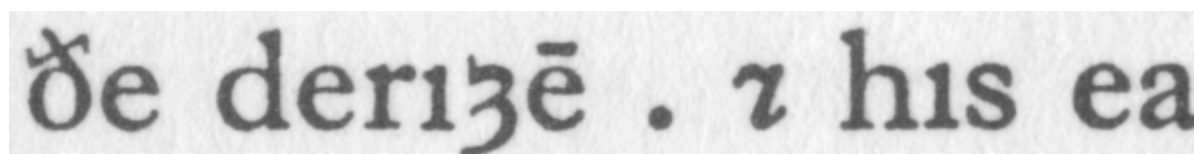

Figure 9. Close-up example of the type cut for ꝛ TIRONIAN ET WITH HOOK as shown in Figure 8. Compare Figure 12; the shape there without the crossbar is also acceptable for this character.
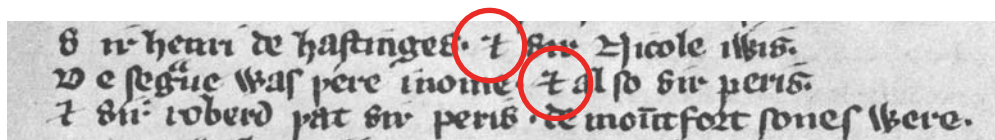
Figure 10. Middle English text, "Chronicle of Robert of Gloucester", written between 1300 and 1325, British Museum Cotton MS Caligula A.xi, fol. 165r. Two examples of ⁊ TIRONIAN ET WITH HOOK AND STROKE can be seen. From Wright 1960, Plate 10.

Figure 11. Transcription of Figure 10. From Wright 1960, Plate 10.
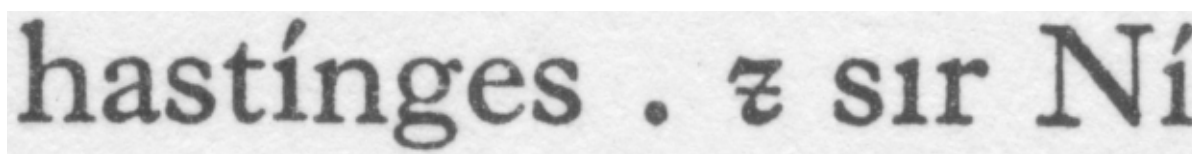

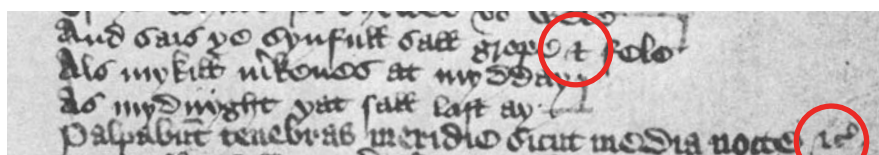**Figure 12.** Close-up example of the type cut for ⁊ as shown in Figure 11.

Figure 13. Middle English text, "The Pricke of Conscience", written 1405 by John Farnelay, British Museum Additional MS 32578, fol. 76r. In the first line ⁊ TIRONIAN ET WITH HOOK AND STROKE can be seen, and in the last line ʒ TIRONIAN ET WITH HOOK is used. From Wright 1960, Plate 16.

Figure 14. Transcription of Figure 13. Here the editor distinguishes both characters in his transcription. From Wright 1960, Plate 16.

Figure 15. Latin biblical text written in 1407 showing an example of ⁊ TIRONIAN ET WITH HOOK AND STROKE: "gl'oꝛ a uigīti ānis ⁊ ſuꝑ oēs" = "*gul*o*rum* a uigi*n*ti a*n*nis sup*er* o*mn*es".
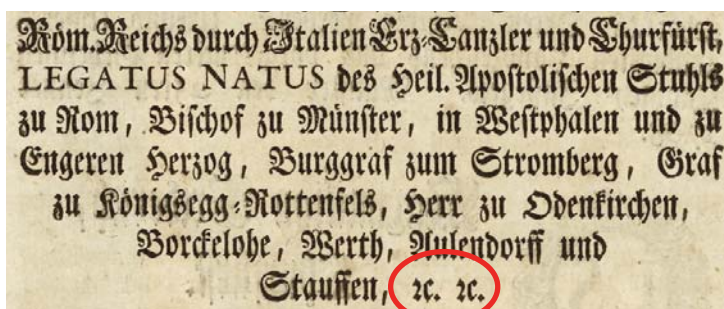

Figure 16. German text dated 1768 showing ʒ in the last line in the abbreviation ʒc. "etc.".
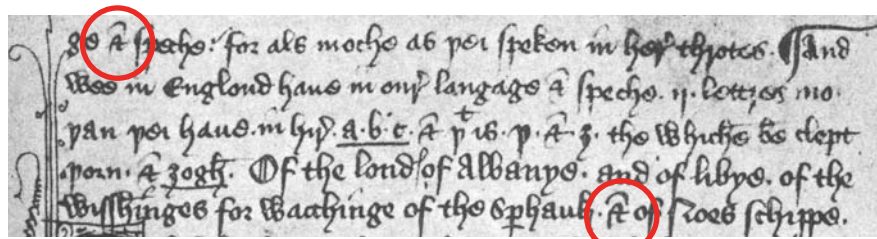
Figure 17. Middle English text, "Travels of Sir John Mandeville", written c. 1410–20, British Museum Cotton MS Titus C.xvi, fol. 60v. Here ⁊ TIRONIAN ET WITH HOOK AND STROKE is written with an inverted breve diacritic (which can be achieved in the usual way with U+0311). As in Figure 13, the headbar of the glyph has acquired a strong downstroke. From Wright 1960, Plate 17.

Figure 18. Transcription of Figure 17. In addition to its literary and historical value, this text is important as it is a source for the names of the letters thorn (þoʀn) and yogh (ʒogh). From Wright 1960, Plate 17.
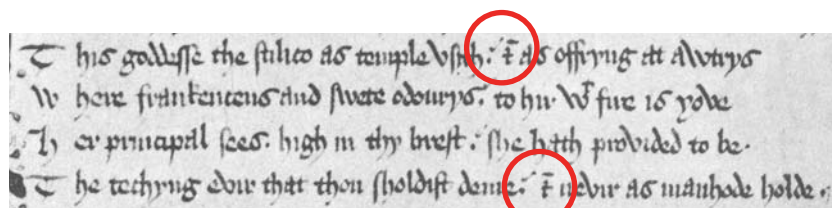
Figure 19. Middle English text, Claudian's "De Consulatu Stiliconis", written in 1445, British Museum Additional MS 11814, fol. 6r. Here ⁊ TIRONIAN ET WITH HOOK AND STROKE is written with a macron (which can be achieved in the usual way with U+0304). Here the headbar of the glyph is greatly truncated. From Wright 1960, Plate 19.

Figure 20. Transcription of Figure 19. From Wright 1960, Plate 19.

Figure 21. English text, "Secrees of Old Philisoffres", written 1469 by William Ebesham, British Museum Lansdowne MS 285, fol. 152r. Here ⁊ TIRONIAN ET WITH HOOK AND STROKE is written with a headbar which is greatly truncated. From Wright 1960, Plate 24.

Figure 22. Transcription of Figure 21. From Wright 1960, Plate 24.

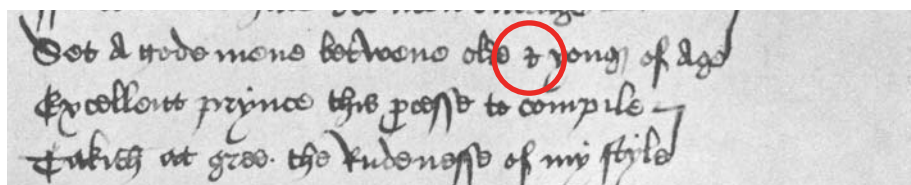|		
&	et	*The & appears in the Print wherever an Abbreviation for* et *appears in the Manuscript.*
ꝛ	and	*In the Copies of some Scotch Records.*
&	eciam	
÷	est	

ꞔ	cer	feꞔit - - - fecerit		ꞔtum - - - certum		
m̄	mer	amꞔietur - amercietur		m̄cator - - mercator		
s̄	ser	s̄vicium - - servicium		s̄vir - - - servir		
t̄	ter	ceꞇa - - - cetera		t̄ram - - - terram		
	sometimes tre	ꞇspasours - trespasours				
ū	uer	fūit - - - fuerit		hūit - - - habuerit		
v̄	ver	estoꝟiū - ·· estoverium		oꝟtes - - - overtes		

ꝑ	per	*As the Prepositions* per *and* pro, *and in forming Words :*				
ꝑ	pro	cepit - - ceperit		ꝑꝑia - - - propria		
		ꝑbū - - - probum		ꝑsone - - - persone		
		aptement - apertement		ꝑfit - - - profit		
		ꝑmisit - - permisit		ꝑmisit - - - promisit		
		ꝑchein - - prochein		ꝑa - - - - persona		
ꝑ	also for par *and* por	ꝑatus - - paratus		ꝑceles - - - parceles		
		tempe - - tempore		corpum - - corporum		
ꝓ	pre	ꝓsentem - presentem		ꝓlaz - - - prelatz		
		ꝓdc̄m - - predictum		ꝓndront - - prendront		
ʒ	tz	fiʒ - - - - - fitz		feʒ - - - fetz		
		establisemenʒ - establisementz		tenaunʒ - tenauntz		

In the Copies of some Records the following Characters occur :

ꝭ	is *or* s	Scottꝭ - - Scottis	*As also the Saxon Characters.*	
		Lordꝭ - - Lords	þ - - th	ȝ - - y

Figure 23. Chart of medievalist characters from the ninth edition (1879) of *The Court-Hand Restored*, a manual of early handwriting whose first edition was published in 1773. Here ⁊ TIRONIAN ET WITH HOOK AND STROKE is represented with a rather clumsy glyph derived from a tilde-like swash and lowercase letter *t*. From Wright 1879.
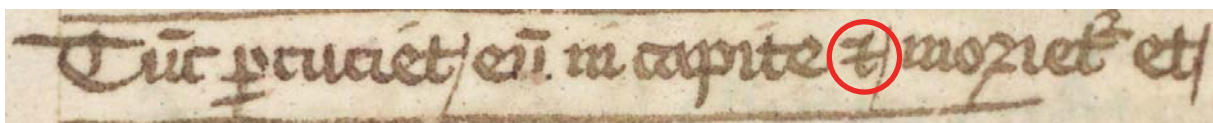

Figure 24. Latin text in the Cornish play, "Origo Mundi", written late 14th century, Bodley MS 791, fol. 6r. The text shows ⁊ LATIN SMALL LETTER TIRONIAN ET WITH HOOK AND STROKE alongside the regular Latin word *et* at the end of the line.
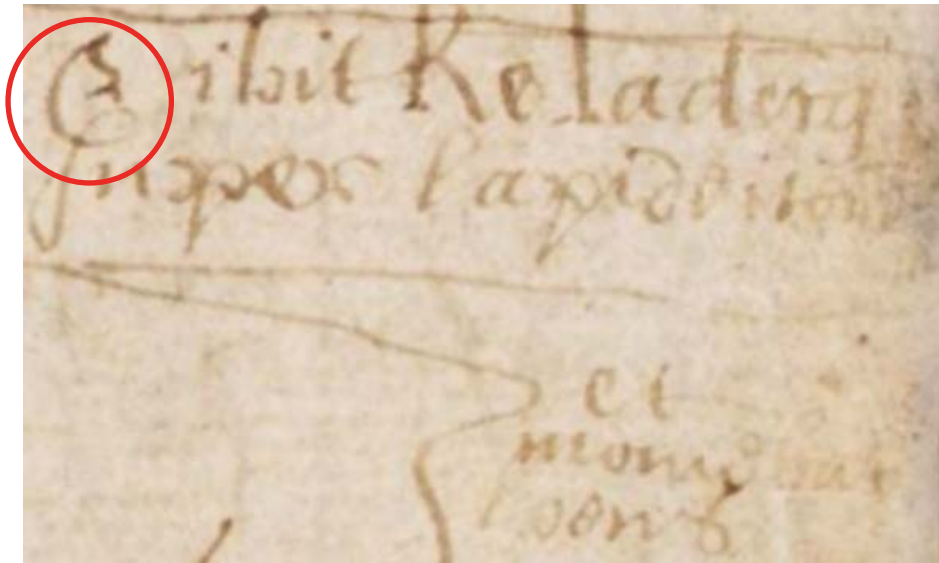
Figure 25. Latin text in the Cornish play, "Bewnans Ke", written late 14th century, Peniarth MS 105B, fol. 1r. The text shows ⅂ ꜱᴍᴀʟʟ ᴄᴀᴘɪᴛᴀʟ ʟᴇᴛᴛᴇʀ ᴛɪʀᴏɴɪᴀɴ ᴇᴛ ᴡɪᴛʜ ʜᴏᴏᴋ with its headbar descending in a leftwards curl. The text reads "⅂ ibit Keladocꝯ ſuper lapidê iterū et manebit ſedens" "*Et ibit Keladocus super lapidem iterum et manebit sedens*" 'And Ke mounts the stone again and will remain seated.'
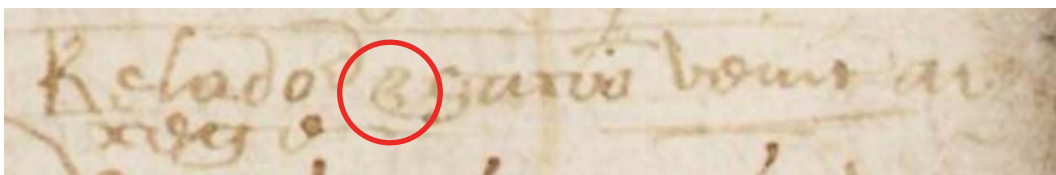
Figure 26. Latin text in the Cornish play, "Origo Mundi", written late 14th century, Peniarth MS 105B, fol. 4r. The text shows ⅂ ʟᴀᴛɪɴ ꜱᴍᴀʟʟ ʟᴇᴛᴛᴇʀ ᴛɪʀᴏɴɪᴀɴ ᴇᴛ ᴡɪᴛʜ ʜᴏᴏᴋ with its headbar descending in a leftwards curl. The text reads: "Keladoꝰ ⅂ garçõ veniūt ad regē" "*Keladocus et garçon veniunt ad regem*" 'Ke and the garçon go to the king'.

Figure 27. Latin text in manuscript about the English royal family written c. 1284, British Library Additional MS 24686, fol. 12r. The text shows ⅂ ʟᴀᴛɪɴ ꜱᴍᴀʟʟ ʟᴇᴛᴛᴇʀ ᴛɪʀᴏɴɪᴀɴ ᴇᴛ ᴡɪᴛʜ ʜᴏᴏᴋ ᴀɴᴅ ꜱᴛʀᴏᴋᴇ with its headbar descending in a short leftwards curl. From Parkes 1993, Plate 21.
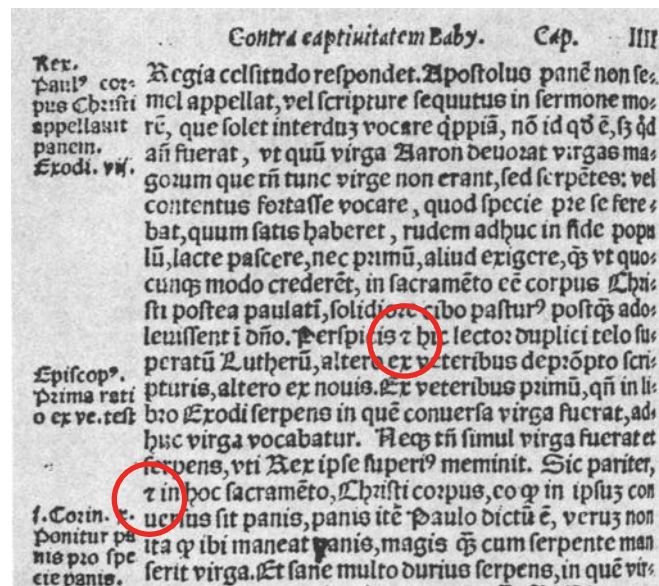
Figure 28. Latin text in J. Fisher's *Defensio regie assertionis contra Babylonicam captiuitatem*, Cologne, written in 1525. The text shows ⁊ LATIN SMALL LETTER TIRONIAN ET WITH HOOK. From Parkes 1993, Plate 36.
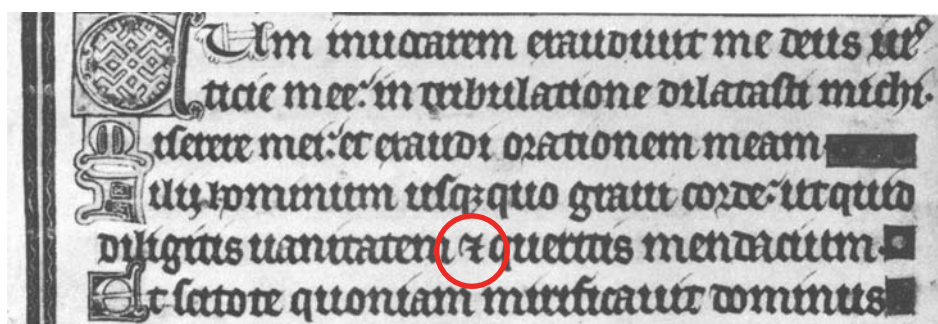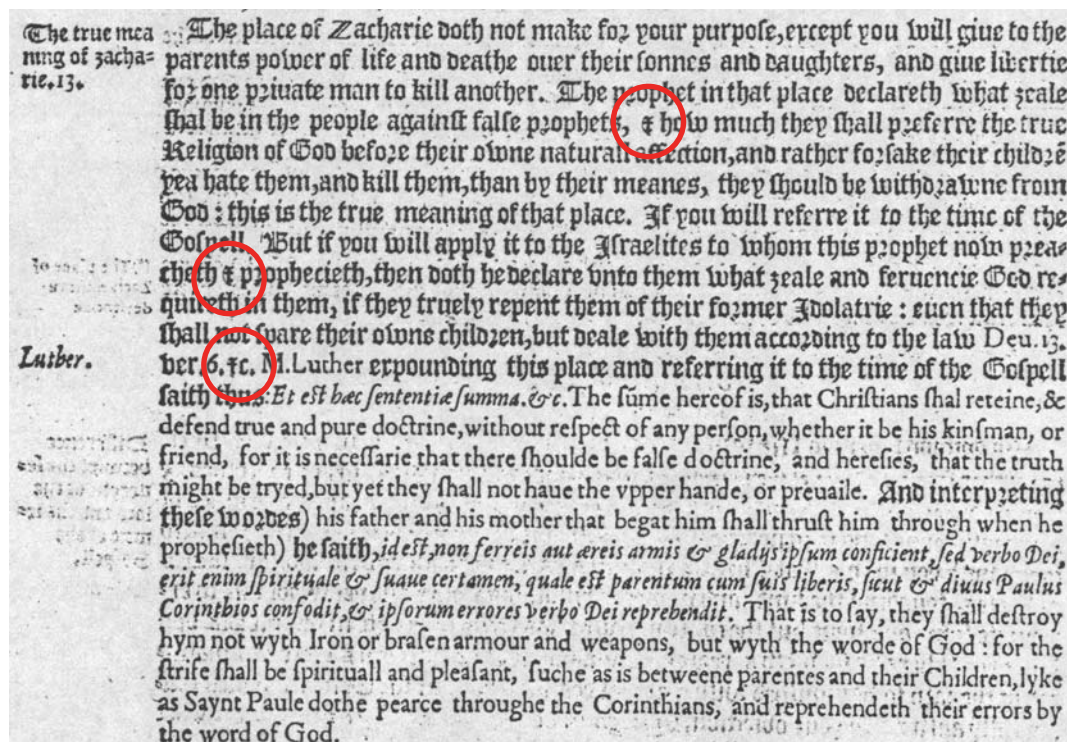

Figure 29. English text in John Whitgift's *The defense of the answere to the admonition against the reply by T[homas] C[artwright]*, London, written in 1574. The text shows ⁊ LATIN SMALL LETTER TIRONIAN ET WITH HOOK AND STROKE with a hairline descending in a leftwards curl from its headline. From Parkes 1993, Plate 37.
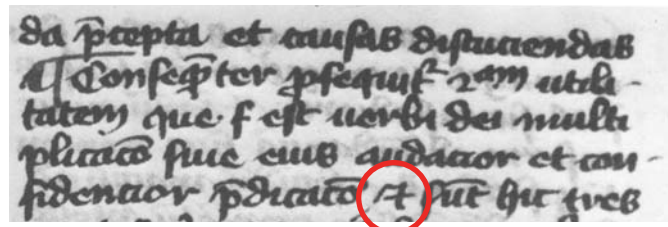
Figure 30. Latin text, a commentary on Paul's letters, written c. 1442, Oxford, Balliol Colleve, MS 30, fol. 119v. The text shows ⁊ LATIN SMALL LETTER TIRONIAN ET WITH HOOK AND STROKE with its headbar descending in a short leftwards stroke. From Parkes 1979, Plate 24 (ii).
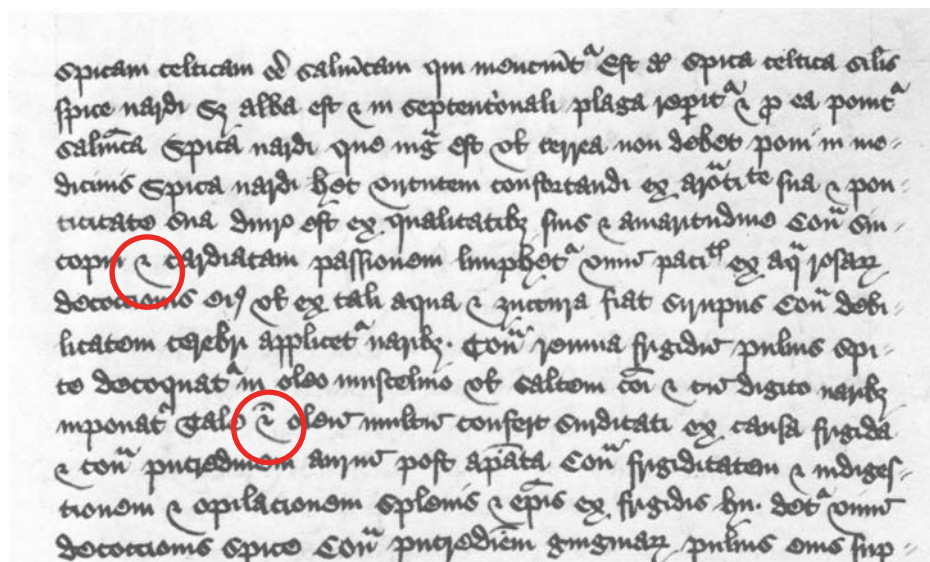

Figure 31. Latin text, a commentary on Paul's letters, written c. 1442, Oxford, Bodleian Library, MS Wood Empt. 15 (*SC*. 8603), fol. 106r. The text shows numerous examples of ⁊ LATIN SMALL LETTER TIRONIAN ET WITH HOOK. nicely drawn with its headbar descending in a long rightwards curve. In the fourth line from the bottom there is a form ⁊̃ combined with U+0353 COMBINING FERMATA, which has the reading *eciam*. From Parkes 1979, Plate 3 (i).

A. Administrative

1. Title

Proposal to add five Tironian Latin letters 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

2017-09-04

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

Latin Extended-D

2. Number of characters in proposal

5.

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.)?

Yes.

2b. If YES, with whom?

Peter Baker, the Anglo-Saxon discussion group ansax-l@listserv.wvu.edu, the Chaucer discussion group chaucer@listserv.uic.edu, Facebook groups for Old and Middle English

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, Anglicists.

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

Common in linguistic texts.

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?

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?