


Characters for Classical Latin

David J. Perry
Version 4, April 18, 2006

Items added or significantly changed after version 3 have red double borders. This will make it easy for those who have read previous versions to locate changes without studying the entire file. 

Introduction

The purpose of this document is to identify *all* characters of interest to those who work with Classical Latin, no matter how rare. Epigraphers will want many of these, but I want to collect any character that is needed in any context. Those that are already available in Unicode will be so identified; those that may be available can be debated; and those that are clearly absent and should be proposed can be proposed; and those that are so rare as to be unencodable will be known.

If you have any suggestions for additional characters or reactions to the suggestions made here, please email David Perry at hospes@scholarsfonts.net. No matter how rare, let's get all possible characters on this list.

Please note that abbreviations constructed with lines above or through existing letters are not considered separate characters except in unusual circumstances, nor are the space-saving ligatures found in Latin inscriptions unless they have a unique grammatical or phonemic function (which they normally don't).

Characters that I think **should be proposed** for inclusion in Unicode are highlighted in gold to make them easy to spot. Characters **that require more discussion** are highlighted in green. **Missing or incorrectly shaped glyphs** are highlighted in purple; these will be fixed in future versions of this document.

Some references are made to “smart font technology.” This refers to OpenType fonts (Windows and Linux, and occasionally on Mac OS) or AAT (Mac) fonts that can do things like position diacritics correctly over a variety of base letters or replace one sequence of characters with a ligature glyph. Such technology is available today, although not supported by many applications; there is every reason to think it will be more widely supported in the future, especially after the release of Windows Vista. If you want more information about this, email me. What users need to know, however, is that they must tell font developers what they need so that fonts can be created.

The Medieval Unicode Font Initiative is a group of medievalists working hard to get their characters into Unicode and to create a standardized solution for those that cannot be encoded. Some of their work overlaps with this document, and when appropriate cross-references are given. See the MUFI website at <http://gandalf.aksis.uib.no/mufi/>.

Thanks to Deborah Anderson of the Script Encoding Initiative, John Bodel of Brown University, and Richard Peever of TLG for helpful suggestions.

List of Classical Latin Characters

I. Letters, including letters with diacritics

A. Long I and other tall letters

?	missing	Unicode does not normally encode letters that are variant forms of items already in the Standard. However, the fact that the long I is intended to represent the phoneme /i:/, distinct from /i/, makes it a candidate for encoding, and it should be proposed . Other letters are occasionally found written in an extra-tall form, but unless they can be shown to have specific phonemic or grammatical values (not the case, as far as I know), they should be treated as glyph variants.
---	---------	--

B. Claudian letters

ⱦ	2132	Two of the Claudian letters (uppercase forms) are already in Unicode. The fact that U+2183 is named ROMAN NUMERAL REVERSED ONE HUNDRED might seem to make it unsuitable as a regular letter, but Michael Everson, author of the proposal for the Claudian letters and expert on Unicode, recommends this use. The third, the half-H character, has been proposed, along with lowercase forms for all three Claudian letters. <u>Tentative</u> Unicode values are U+214E TURNED SMALL F, U+2184 LATIN SMALL LETTER REVERSED C, and U+2C75/76 LATIN CAPITAL/SMALL LETTER HALF H.
ↀ	2183	
Ɔ	proposed	

C. Name abbreviations

ↀ	2183	The reversed C sometimes found for Gaia should be represented by this character; see above under Claudian letters for justification of using this “numeral” as a regular letter.
ⱦ	missing	The five-legged variant of M used for Manius is missing and should be proposed . The MUF I spec has it in the PUA at F227 with the name MARIAN M, which I don’t understand, but cites a German work on Latin inscription. More research needed about that..

D. Ligatures

ZWJ	200D	Unicode does not encode ligatures, except when they can be shown to have special phonemic values (not true for ancient Roman usage, where ligatures are mainly space-saving devices). Encode them by entering the first letter, then a zero width joiner, then the second letter. Smart fonts can replace such sequences with the appropriate glyph.
-----	------	--

E. Apex and other markers of vowel length

Á	00C1	For the apex, use the acute accent. A, E, I, and O exist in precomposed form; V-apex should be encoded with the combining acute (U+0301). If a PUA codepoint for the latter is needed until combining marks are handled better, use E33A, as found in the Titus font and the MUFI spec. Fonts may contain glyph variants if it is desired to use a different shape for the apex, such as the curved hook form. Using other diacritics such as xx for alternate apex shapes is not recommended.
É	00C9	
Í	00CD	
Ó	00D3	
Ā/ā	0100/1	If it is desired to use the modern convention of a macron to indicate length, all six vowels with macron are present in precomposed form in Unicode (given at left for easy reference). The 19 th century convention of using a circumflex for long vowels can be represented, if needed, with the precomposed circumflex vowels. To mark a vowel specifically as short, the precomposed vowel + breve combinations are recommended. Y/y plus breve is not encoded, however, so Y/y followed by U+0306 COMBINING BREVE would be used.
Ē/ē	0112/3	
Ī/ī	012A/B	
Ō/ō	014C/D	
Ū/ū	016A/B	
Ȳ/ȳ	0232/3	

F. Sicilicus

◌ [◌]	0357	This relatively rare diacritic (representing a doubled consonant, esp. over V in words such as SERVVS=servus) can be encoded using U+0309 COMBINING HOOK ABOVE or 0357 COMBINING HALF RING ABOVE; the latter seems the better choice to me since the former is intended as a tone mark in Vietnamese and since the shape is closer to that used by the Romans.
----------------	------	--

G. Overline for general abbreviations

◌ [—]	0305	U+0305 COMBINING OVERLINE is specifically designed to connect visually to form a continuous line, which is usually what we find in Latin inscriptions. If adjacent letters with non-connecting overline are desired, use a macron, either the precomposed forms when they exist, or the COMBINING MACRON U+0304 after a base letter. Smart font technology is needed to position the combining marks correctly, both horizontally and vertically.
----------------	------	---

H. Line through for general abbreviations

◌ [⊖]	0336	The COMBINING LONG STROKE OVERLAY is the appropriate character for such abbreviations. Smart font technology is needed to position the overlay at the right height and to supply a glyph with an overlay of the proper length. There is also a COMBINING SHORT STROKE OVERLAY at U+0335. I suggest standardizing on one overlay for abbreviations and using font technology to get the desired appearance.
----------------	------	--

II. Numerals

Roman numerals should normally be represented by the standard Roman letters I, V, X, L, D, C, and M. The Roman numeral forms in Unicode that duplicate the standard Latin letters (U+2160–217F) were put there for use in East Asian text (often printed vertically) and should not be used in Latin script.

◌̄	0305	Roman numerals are barred above to indicate multiplication by 1,000, when used as a prefix on words (IIviri = duumviri), when used as ordinal adverbs, and sometimes for other reasons. In all these cases, use U+0305 COMBINING OVERLINE; see I.G above
	007C	A bar above and a vertical line on either side of a numeral indicate multiplication by 100,000. Use U+007C VERTICAL BAR for this purpose. Smart font technology may be needed to get the corners of the overline and the vertical bar lined up right.
ϸ	missing	This form of the numeral six is missing and should be proposed . It is found most often in Christian inscriptions.
Ɔ	glyph var.	The Romans wrote the numeral for 500 with a horizontal bar. I suggest treating this as a glyph variant of the letter D , since standard Latin letters are normally used for Roman numerals. This also makes it easy to search for (users don't have to enter a special codepoint). It could, however, be encoded as D followed by 0336, COMBINING LONG STROKE OVERLAY.
↴, ↵, ⊥	glyph var.	For similar reasons, I suggest treating the early forms of the numeral fifty (downward pointing arrow, inverted T, anchor shape) as glyph variants of L . I know that this is wrong from a chronological viewpoint, but I think it is the most practical way to proceed.
∞	glyph var.?	The Romans never wrote 1000 as the letter M except in some abbreviations (MP = <i>milia passuum</i> , e.g.). The numeral originally looked like a circle with a vertical line through its center, later rather like a modern infinity sign. As suggested above in the case of 500, one might argue for treating this as a glyph variant of M. However, the shape is so totally different that a new character might be proposed. But there would then be two different ways of representing the Roman numeral 1000 in Unicode. A better way might be to handle it as a glyph variant of U+2180 (see below), from which the infinity-shape may have evolved. The mathematical infinity sign in Unicode U+221E is not well suited for use as a Roman numeral.
ↀ	2180	ROMAN NUMERAL ONE THOUSAND
ↁ	2181	ROMAN NUMERAL FIVE THOUSAND
ↂ	2182	ROMAN NUMERAL TEN THOUSAND

↷	2183	ROMAN NUMERAL REVERSED ONE HUNDRED Unicode states that the letter C can be used with ↷ 2183 and the letter I to represent larger values; for example, GORDON. Alternate glyphs may be needed for good appearance in such situations.
---	------	---

III. Units of Weight




Tables of weights are found in di Stefano Manzella p. 309 and *OCD*³ p. . They agree except with regard to the sextula; see below.

I	0049	The libra (pound) is represented by the Roman numeral I, for which use the letter I.
-, =	missing	The uncia is shown by a horizontal line and the sextans by two lines, one on top of the other. The shape is similar to the Aegean numbers one and two, U+10110 - and 10111 =, except that the lines are longer. The uncia has glyph variants including a dot (common on Republican coins), a breve-shape, and a lazy-S. The sextans has a glyph variant z (probably the two lines formed without lifting the pen) or Z. See Capelli p. 411. Because of the glyph variants, it is probably not appropriate to unify the uncia and sextans with the two Aegean numbers, so they should be proposed.
S	0053	The semis (1/2 as or 6 unciae) is represented by the letter S, which is also used in other situations as an abbreviation for 1/2. Note that all the fractions from the deunx down to the quadrans can be built by combining the letter S, the uncia, and/or the sextans; see table. It seems reasonable to make them by combining these glyphs. However, the Aegean numbers are encoded individually rather than being built up from smaller units; is that parallel relevant here?
Σ	03A3	The semuncia is denoted by a Greek Sigma, with glyph variants € and £ without the crossbar. I believe that these glyph variants may in fact be cursive forms of the Sigma, but they are not found in Greek texts; so the unification with Sigma is not appropriate, and the semuncia should be proposed.
↷		The sicilicus (1/48 of an as or 1/4 of an uncia, and not the same as the diacritic for doubled consonants) is marked with a sign that appears similar to U+10140, GREEK ACROPHONIC ATTIC ONE QUARTER; is this an appropriate unification? U+03FD GREEK CAPITAL REVERSED LUNATE SIGMA SYMBOL and U+2183 ROMAN NUMERAL REVERSED ONE HUNDRED do not have the right appearance (arms too long).

ℷ	missing	The sextula (1/72 of an as or 1/6 of an uncia), represented by a reversed S glyph, is missing and should be encoded . Binae sextulae is shown by two of these reversed S glyphs and dimidia sextula by the same reversed S or its glyph variant ℷ.
Ϸ	2108	Unicode contains a symbol for the scrupulum at U+2108. This apparently was added to represent a medical symbol. Since it has the correct shape, is there any reason not to use it for the ancient Roman scripulum (1/288 of an as or 1/24 of an uncia)? The Roman scripulum has a glyph variant Ϸ.
	missing	The siliqua (1/1728 of an as or 1/144 of an uncia) is missing and should be encoded .

IV. Monetary Signs

NB: currency property

⌘	missing	The denarius sign is missing and should be proposed . There are hardly any glyph variants; the Vindolanda tablets have ⌘, which is probably just scribal haste rather than a true variant.
Ϛ	missing	The quinarius sign is missing. It is found less frequently than the denarius or sestertius, but should be included for the sake of completeness .
HS	missing	The sestertius sign is missing and should be proposed . Glyph variants include HS, H, S, SS others?
??	missing	Cagnat and Ricci give signs for the dupondius,  /  and  respectively. I have not seen these elsewhere. I would like to see one in an inscription, or some other confirmation, before proposing the dupondius, although it should probably be proposed for completeness.
	2758??	Cagnat and Ricci give the <i>as</i> as a slanted line with horizontal bar, while Sandys says that it is normally a simple vertical line. U+007C has the right shape, although in some fonts it descends below the baseline; but it also has the math property (Sm). 01C0 LATIN LETTER DENTAL CLICK has the letter property (Lo); 05C0 HEBREW PUNCTUATION PASEQ is a punctuation character (Po) specific to the Hebrew script; 2223 DIVIDES is also a math symbol; and 2758 LIGHT VERTICAL BAR is Symbol, other. The latter may be the closest to what we need, or should a separate character be proposed? The Vindolanda tablets have a glyph variant 7, a slanted line with a tick, taller than adjacent Roman numeral one. Given that the handwriting on the tablets usually slants to the right somewhat, it might be best to regard this simply as a tall vertical line with a tick to distinguish it as an abbreviation. I am leaning toward proposing the <i>as</i> separately, based on the glyph variants.

Note that some Republican coins use the same weight symbols discussed above (I for one *as*, S for semis, and dots rather than lines for the uncia and its multiples).

IV. Other characters

⌞	missing	The centurial sign can stand for several forms (centurio, centurioni, centurione, centuriae, others?). It should be proposed for Unicode. Glyph variants include a reversed C (sometimes barred), the letter O, an angle bracket shape, a 7-shape, a vertical squiggle and the letter Z. The 7-shape is the most common and so might be the reference glyph.
θ	03B8	The Greek letter theta is used to indicate that a person is deceased.
v	0076	v = vivit, normally smaller and raised to the left of a person's name in lists. Could this be encoded separately as a symbol??
Ϝ Ϛ	03DC 03DA	Some Roman coins minted in the East, with legends in Latin, use the Greek digamma or stigma with the numeric value of six. Use the appropriate Greek characters in such cases, with glyph variants as needed (e.g., sometimes the Digamma looks more like a Gamma).
Ϟ	2627	CHI RHO

V. Punctuation

·	00B7??	MIDDLE DOT, U+00B7, is the character that I have suggested in the past for the interpunct. Another option is U+2022 BULLET, which has a larger diameter than the middle dot. Both of these have the punctuation property, are found in many fonts, and have approximately the right shape. Another issue is line breaking; apps should break lines, if needed, after an interpunct even if no space character is entered. It doesn't appear that this happens—definitely not in Word XP. NEED TO CONFIRM.
⬮	2023??	
		In inscriptions, aside from the dot, one finds horizontal wedge-shapes, triangles, and (in Republican texts) a square; all made with one stroke of a punch. There are also hederæ and star-shaped forms (latter made with two overlapping strokes of a triangular punch). U+2023 TRIANGULAR BULLET has an appropriate shape, but exhibits the same line breaking behavior as the two round characters.
		Assuming that lines will not break after the three characters discussed above, we should propose a new character, probably with a wedge-shaped reference, from which the other shapes could be treated as glyph variants. I think that this is better than proposing all the different shapes separately, but am willing to listen to arguments about this.

¶	glyph var.	<p>IGNORE THIS CHARACTER until I can come back and fix it. The paragraphus is used to mark textual divisions. Rather than proposing a new character, I suggest treating this as a variant (actually, several variants) of U+204B REVERSED PILCROW SIGN ¶ or U+00B6 PILCROW SIGN ¶. The latter is the usual division marker, but the former has a shape more similar to that found in MSS. U+2E0F PARAGRAPHOS is probably not suitable since it has the wrong shape is specifically identified as an ancient Greek textual mark. The MUFI recommendation has the medieval paragraphus in the PUA at F1E1. I would like information about this character in classical Latin; Capelli lists a P with right foot shape and nothing else.</p>
•	2767	<p>ROTATED FLORAL HEART BULLET is specified in Unicode as a heder. U+2619 REVERSED ROTATED FLORAL HEART BULLET • is also available. However, these shapes are not useful for classical Latin because a) the stems are horizontal not vertical and b) they have the symbol rather than punctuation properties. They are traditional shapes for printers' ornaments in western typography. See above for more about interpuncts.</p>
𐌸	missing	<p>The palm branch, occasionally found to mark divisions of text and as a decoration on coins, is missing and should be encoded. MUFI spec has it in the PUA at U+F1DB.</p>

VI. Modern Editorial Signs

*	002A	
[]	005B/005D	
{ }	007B/007D	
< >	003C/003E	
†	2020	
‡	2021	
◌̣	0323	Unicode contains a large number of precomposed letters with underdot in the Latin Extended Additional range, including all the Latin alphabet. Any additional characters with underdot that may be required should use combining underdot, U+0323, in which case smart font technology will be needed for good appearance.
< >	2329/2A	LEFT-POINTING ANGLE BRACKET / RIGHT-POINTING ANGLE BRACKET. Do not use the similar-looking U+3008/09, which are specific to East Asian text and are normally double-width, or U+27E8/E9, which are for mathematics.
[[]]	27E6/E7	
⸀ ?	2E00/missing	The upper left bracket was included in the set of sigla used in New Testament studies, but the upper right bracket is missing and should be proposed .
? ?	missing	Lower brackets are missing and should be proposed .
◌̣◌̣	0361	COMBINING DOUBLE INVERTED BREVE may be used over two letters to indicate a ligature. If it is preferred to use a circumflex over the first letter of a ligature, the many precomposed circumflex forms may be used, or a base letter followed by COMBINING CIRCUMFLEX.
◌̣^	0302	
⊂ ⊃	missing	The sideways-pointing U shaped brackets used in CIL to denote Claudian letters, inverted letters, or images are missing and should be proposed . The MUFIspec has them in the PUA at F1DC/DD. The SUBSET OF and SUPERSET OF characters (U+2282/83) are mathematical operators and so should not be used; in addition the legs of the glyphs are too short.
(())	missing	Some editors use double parentheses for similar purposes similar to the sideways U brackets. These also are missing and should be

| **proposed.** The MUFI spec has them in the PUA at F1DE/DF.[†]

[†] The MUFI group is preparing a proposal for medievalist and Latinist punctuation, the first draft of which contains both the sideways U brackets and the double parentheses.