

UC Berkeley

Proposals from the Script Encoding Initiative

Title

Medieval Latin Character Recommendations

Permalink

<https://escholarship.org/uc/item/8ns5m5pz>

Author

Perry, David

Publication Date

2020-08-11

Peer reviewed

Medieval Latin Character Recommendations

For Works in

The Digital Library of Late Antique Texts

David J. Perry

Version 1, updated August 14, 2020

Background about Texts in Classical, Late, and Medieval Latin

Unicode contains many characters needed by those who work with ancient and medieval texts in Latin. As a starting point, you can consult [a reference list](#) of all the characters needed for classical Latin; many of them continued in use through Late Antiquity and into the Middle Ages.

About This Document

This document is a response to a list submitted by the editors of the [Digital Library of Late Antique Texts](#), containing letters, numbers, or symbols for which they were unsure which Unicode characters to use. It was prepared with the assistance of Dr. Deborah Anderson of the Script Encoding Initiative at the University of California Berkeley. It was also made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project (NEH PR-253360-17). The Universal Scripts Project is part of the Script Encoding Initiative at UC Berkeley. Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

This file is a work in progress, since editors are still adding new texts to the Library. Editors should consult this document to see if the letters, numbers, or symbols they need are already included in the attached lists with recommended Unicode characters; if not, they should contact David Perry. This document also includes two sets of characters that require more information before a final recommendation can be made; those items that are highlighted in green and those that are contained in section V.

As new requests come in, the author of this document will add new entries, with the recommended Unicode characters to represent them. In some instances, there may be no existing Unicode characters that are appropriate. In such a case, a Unicode proposal will need to be made, with directions as outlined in the [“Submitting Character Proposals”](#) page. Deborah Anderson can assist on such a proposal.

Submitting Characters

For characters that are not included in this document or for which more information is requested, editors should send:

- A scan showing the shape
- The meaning or value of the letter, number, or symbol (if known, or even an educated guess); this is important.
- Any additional information that is available, particularly when the shape of the character seems unrelated to its value. See the discussions below for examples of the kind of information that is helpful.

Send the above information to David Perry at hospes02@scholarsfonts.net, with a copy to Deborah Anderson at dwanders@berkeley.edu or dwanders@sonic.net.

General Considerations

Users should understand that the best representation by Unicode characters for a letter, number, or symbol is that which presents its meaning in the most accurate (i.e., least ambiguous) way. This is *not* necessarily the same as reproducing the shape found in any given source. For example, what looks like a capital F is found representing four oboli. This should be encoded using U+1017F GREEK FOUR OBOLS SIGN not U+0046 LATIN CAPITAL LETTER F since the latter does not specifically indicate the monetary unit involved. (See below for ways to enable users to see a Latin capital F rather than the Greek symbol ₴.)

If a character is found in various texts with different appearance, editors should be consistent about using the same encoding for that character.

Displaying Alternate Shapes

For those who are not familiar with the basic glyph versus character issues, [this technical note](#) may be helpful. Another, more detailed resource on this issue is [here](#); medievalists can skip the part about right to left directionality, which is not an issue for them.

The bottom line for projects such as the Digital Library of Late Antique Texts is that it is possible, if desired, to reproduce the shape found in a MS instead of the default shape found in a font. This is particularly true since HTML and CSS offer good support for the use of OpenType features. Whether doing so is worth the effort and potential expense is something that project editors must decide. Using the best, most appropriate encoding should be the first step for a project that aims to make these texts available on line for the long term.

Miscellaneous Notes

References to Cappelli are to the 1979 edition.

Part I. Roman Weights and Measures

Note that the Romans used many of the items on this page for both weights/coin values and measures. For example, the denarius is a common coin symbol and also occasionally found as an apothecary's weight. Since the ancients themselves used the same symbols in two ways, I do not see any need to encode the measurement symbols separately. If there are situations where it cannot be determined which is intended simply by reading the text, then additional characters might be proposed.

Variants marked with a superscript italic 1 are some that did not surface when the original proposal for Roman weights and measures in Unicode was prepared. These variants are found in a list of characters prepared by Maurizio Lana and his colleagues for the [Digital Library of Late-Antique Texts](#). Orange shading indicates an approximate glyph; see the scans in the PDF file for exact medieval shapes.

The table below recommends U+0053 LATIN CAPITAL LETTER S for the semis. This would be appropriate when transcribing, e.g., a Roman inscription that is written in monumental capitals. U+0073 LATIN SMALL LETTER S could also be used and might be more appropriate when transcribing, e.g., a manuscript written in uncials. Users could search successfully using either form (unless 'match case' was explicitly chosen).

If a combination such as "S==" doesn't look appropriate or attractive, font makers can provide alternate glyphs (e.g., with the bars farther apart vertically).

NAME	UNICODE	GLYPH	VARIANTS
As	1019A	𐌰	/, l, 𐌰 ¹
Deunx	0053 10190 10190 10191	S==	S= ¹ , S=/ ¹ , SZZ/ ¹
Dextans	0053 10190 10190	S==	SZZ ¹
Dodrans	0053 10190 10191	S=-	S= ¹ , S=/ ¹ , sz/ ¹
Bes	0053 10190	S=	sz ¹
Septunx	0053 10191	S-	S/ ¹
Semis	0053	S	
Quincunx	10190 10190 10191	==	= ¹ , ZZ/ ¹
Triens	10190 10190	==	:: ¹ , ZZ
Quadrans	10190 10191	=-	:: ¹ , =~ ¹ , Z/ ¹
Sextans	10190	=	z, Z
Sescuncia	10192 10191	𐌱-	𐌱 ¹ , Σ ⁻¹ , 𐌱 ¹ , -𐌱 ¹ , N𐌱 ¹
Uncia	10191	-	•, u, ∞, 𐌵, (𐌵 ¹ , ~ ¹ , / ¹
Semuncia	10192	𐌱	Є, Σ, 𐌱 ¹ , s ¹ , 𐌱 ¹ , 𐌱 ¹
Binae sextulae / Duella	10193 10193	22	ll, 𐌶 ¹ , 𐌶𐌶 ¹
Sicilicus	10140	𐌶	
Sextula	10193	2	𐌶, ζ, ¹
Dimidia sextula	10194	𐌶	𐌶 ¹ , ψ ¹
Scripulum	2108	𐌶	𐌶 ¹ , 𐌶 ¹ , S ¹
Siliqua	10195))	- ¹ , s ¹

𐌱: = Sigma with round upper left corner;
 𐌱: = flipped Є
 Є

Part II. Other Characters

When Multiple Characters Are Available

If there are two or more encoded alternates, as with Pondo, you may choose to match the appearance of the source text or to settle on one for use throughout. The latter might be preferable if you wish to make it easier for users to search for text containing the characters in question, since they would not have to perform multiple searches. But using both forms is certainly a correct encoding, since both are in Unicode.

Using Greek Characters in Latin Text

In a few cases characters encoded for use in Greek are recommended. As a general rule Unicode does not recommend mixing scripts, but there are exceptions. In this case, even though the text is in Latin, the reference is to Greek monetary units (drachma and obol), and the characters were encoded specifically for those units. Scribes clearly changed the original Greek forms to look like Latin letters; glyph variants can be used if one desires to display the medieval Latin form rather than the original Greek shape. Using the specific monetary characters (as opposed to a Latin E, F, etc.) eliminates any possible ambiguity about the intended meanings.

Use of Combining Letters

For proper display of characters encoded with combining letters, fonts should include these combinations in the Mark to Base OpenType feature. Cardo already has this for M + ◌ so the Modius sign below is shown correctly while Mina is not.

Note: in the table and the following discussion, **green shading** indicates questions that should be addressed before a final encoding recommendation is made.

Pondo	A753 A751	ρ ρ	These are lowercase; uppercase forms are also encoded if more appropriate (U+A750 & A752).
Acetabulum	039E <super>03BF	Ξο	See discussion below.
Argenteus	0054	Τ	See discussion below.
Cerates	0292	ζ	See discussion below.
Ceratium	039A <super>03B5	Κε	See discussion below.
Chalcus	004F 0072	Ορ	See discussion below.
Choenix	03A7 <super>03BF	Χο	See discussion below.
Congius	01B5	Ζ	See discussion below. A lowercase form U+01B6 ζ is also encoded.
Cotyle	004C 0053	Λς	See discussion below.
Cyathus	039A <super>03C5	Κυ	See discussion below. Variants: ΤΙ, ΚΥ, wavy line with dots
Denarius	10196	Χ	
Drachma (dragma)	1017B	δ	δ, Χ, Z, ζ, drg, < w/out dot
Dupondius	10199	Η	
Hemina	??	ϸ	See discussion below; propose? Κ°, Ι
Libella	2014	—	See discussion below.
Libra	00A3	℥	See discussion below. ^, /N
Mina	004D 1DE1	Μ̅	
Modius	004D 0366	Μ̅	<insular>m. See discussion below.

Obolus	1017C	~	—,
Dimidius obolus	10140	??	z, τ ; see discussion below.
Duo oboli	1017D	≈	z, =, —, >, > with closed end, r
Tres oboli	1017E	ƒ	T
Quattuor oboli	1017F	f	F
Quinque oboli	10180	ƒ	E, €
Quadrantal	A757 <sup>006C	q ^l	
Quartarius	0069 004D	iM	See discussion below.
Quinarius	10197	∀	
Semodius	004D 0366 0073	Ṁs	s could be raised.
Sestertius	10198	HS	
Sextarius	0186	⊃	Σ, Σ⊃, Ξ ^ε , ζ, s, \$ bar raised diagonally
Siclosescli	0186 <sup>03C5	⊃ ^u	See discussion below.
Siliqua	10195	»	See Part I for variants.
Singula	03A3	Σ	Sigma with round upper left shoulder.
Solidus	004E	N	See discussion below.
Dimidius solidus	0049 0042	IB	See discussion below. Var. I
Talentum	03A4 <sup>039B	T ^Λ	See discussion below. Var. TL
Teruncius	0054	T	
Tremissis	0048	H	See discussion below.
Victoriatus	0245	Λ	Lowercase is 028C if needed.
Vncia	10184	ƒ	
Vrna	A757 <sup>0073	q ^s	

Part III. Abbreviations and Multipliers

General abbreviations with lines over or through letters

Latin texts often make use of abbreviations indicated by a line over the letters or through the letters.

- For overbars, place U+0305 COMBINING OVERLINE after each letter that makes up the abbreviation. Note that this character is specifically intended for bars that connect on either side to form a continuous line. It can be used with Greek letters also since combining marks such as U+0305 are not script-specific.
- For bars through the letters, place either U+0335 COMBINING SHORT STROKE OVERLAY or U+0336 COMBINING LONG STROKE OVERLAY after each letter. The latter may be preferable.


Bars over numerals / Multipliers

Roman numerals appear with overbars for various reasons, most commonly to indicate ordinal adverbs or multiplication by 1,000. Here again U+0336 COMBINING LONG STROKE OVERLAY is the appropriate character to use.

Multiplication by 100,000 is indicated by a numeral surrounded by lines on either side and the top: It is possible to encode this with U+0070 VERTICAL BAR and a combining overline, as follows: 0070 <numeral> 0336 0070. Such an encoding is clumsy and will be difficult to get to appear well.

X

A proposal was made in 2004 to encode a combining mark for Roman numerals multiplied by 100,000 under the name CANOPY. It was rejected by the Unicode Technical Committee in favor of markup.¹ It might be possible to revisit this proposal; otherwise, the combination of U+0070 and 0336 or some sort of markup seems the only option.

There is another multiplier, with an example from Marcellus Empiricus, *De medicamentis Praef.*  Following the numeral with U+2032 PRIME seems a reasonable encoding. It would be good to confirm that this is a 100,000 multiplier. I have never seen it before and would like any other information that is available about it since it seems rare.

IV. Discussion of Specific Characters

Some of the notes below were placed here for my own information when I was not familiar with a character and wanted to know more about it. In some cases there are questions that need answering before we can finalize a recommended encoding; these are marked with green shading.

- Acetabulum: = 1/48 of a congius, liquid or dry measure. Same as Greek ὀξύβαφον (sometimes misspelled ὀξόβαφον). Name is the same as the vinegar cup often put out on dining tables (and the unit, presumably, was more or less what typical vinegar cup held). Unit is frequent in Pliny the Elder. Could the Ξ^o shape possibly have to do with the Greek equivalent?
- Argenteus: What the T glyph has to do with this is unknown. Very strange.
- Cerates: This is an IPA letter also used as a dram sign; the latter use makes it acceptable here, even though it's generally best to reserve IPA characters for specific pronunciations etc. Cappelli (p. 412) shows a Z with rounded bottom with dot on either side as sign for cerates, 8th century.
- Ceratium: Latin **cerātium**, *carob*, from Ancient Greek κεράτιον, *little horn, carob*. So definitely a Greek loanword in origin. As a unit, carat = Roman siliqua 1/728 of a lb. Because some of the other units borrowed from Greek, such as the choenix and cyathus, made use of Greek letters in their abbreviations/symbols, I recommend the Greek letters here. Also relevant is the fact that the Romans hardly ever used the letter K, which points to a Kappa here, and the shape shown in the scan is more like an epsilon than a roman e. It occurs in Greek texts as a slashed Kappa and raised epsilon.
- Chalcus: Greek χαλκός, bronze, copper, money; seems that there was a unit of money by this name in some Greek states. 'Or' seemed a very strange sign until I remembered the Latin **orichalcum**, *copper ore, brass* (= Gr ὀρείχαλκος). Having made this connection I am comfortable with Or as the encoding.
- Choenix: See the comment under Ceratium for rationale of using Greek letters.
- Congius: The slashed Z looks right, but again, this seems quite strange. Origin??
- Cotyle: the TLG encodes this as Kyathos base (Kappa with diagonal slash through the lower leg) + <super>omicron. What does 'LS' have to do with this? The 'LS' looks clear but I would be happier if I knew the origin.
Cotyla (cotyle) = 1/12 of a congius, same as the hemina (liquid or dry measure).
- Cyathus: Another strange one, 'TI' – any explanation? Unless this is much more common than the others, I suggest going with Greek letters here as with some of the other measurements we have. A Kyathos base with upsilon is the Greek abbreviation.
- Hemina: What is the origin of this? It looks similar to U+2185 ROMAN NUMERAL SIX LATE FORM. U+2185 is not appropriate to use, since the hemina is 1/12 of a congius; if it were 1/6 one

¹ Thanks to Dr. Deborah Anderson for researching this.

might make an argument. The K^o form is explained by the equivalence of the kotyle and hemina. The ‘I’ glyph is mysterious. This one might be proposed for Unicode.

- Libella: A libella = an as, 1/10 of a denarius.
- Libra: see email to Maurizio 7/14.
- Modius: this looks like an insular letter M (or m? not sure from the scan) with a small ‘o’ on top. I have therefore suggested 004D 0366; let me know if the small letter is more appropriate. If this is found in MSS written in other scripts (non-insular style) where the insular M is retained, one could make the argument that this took on an identity of its own as a symbol and should therefore be encoded separately. What we decide about Modius will also affect the encoding of the Semodius and Quartarius.
- Dimidius obolus: If we go with the encoded Greek characters for one, two, etc. oboli (see discussion above), then it makes sense to use U+10140 GREEK ONE HALF OBOL SIGN for this character even though the glyph shape ⲟ is not appropriate. There is also U+10175 GREEK ONE HALF SIGN; is it possible that its shape ⲗ is related to the ‘z’ variant of this character? Cappelli (p. 405) shows ‘Z’ as a medieval sign for obolus dimidius. It’s possible that what looks a squished T in the scan is a bar with a dot below; cf ⲗ for one obol.
- Quartarius: The encoding of the second part of this character depends on what we decide about the modius (see above). For the first part, an ‘i’ seems the most obvious choice. If there is information about what the ‘i’ stands for, that might help; is this connected in any way with the value 1/24? U+0271 ⲗ LATIN SMALL LETTER IOTA is an IPA character and so best avoided; if it is desired to encode this shape specifically, better to use a normal lowercase Greek iota 03B9 or perhaps U+0131 LATIN SMALL LETTER DOTLESS I.
- Sextarius: U+1086 appears to be the correct encoding, but again, some explanation or history would be helpful. Value is 1/6 of a congius. This is the same as the Greek xestes, whose abbreviation ⲗ is found in the TLG; the ⲗ^{E} variant is consistent with this. For the slashed s, use U+0335, COMBINING SHORT STROKE OVERLAY.
- Siclosescli: I could not find any information at all about this character, not even its value. The suggested encoding looks correct, but I really would like some information.
- Solidus: Lewis & Short says this is a gold coin of the Empire, originally called aureus, worth 25 denarii, later debased. Could the N possibly be connected with ‘nummi’?? Cappelli shows N as an abbrev. for soldi (p. 411).
- Dimidius solidus: Very strange; would really like some info about where this sign came from.
- Talentum: If we’re going to use Lambda as the second part, I suggest using Tau rather than Latin T for the first half. Another option would be to use U+1017A ⲗ , which specifically encodes the Greek talent. (Is the shape ⲗ possibly a Tau Lambda ligature in origin?)
- Teruncius: 3/12 of an as, a quarter as.
- Tremissis: In Lewis & Short as **tremis**, **tremissis**. A gold coin worth 1/3 of an aureus. Any info available to explain the H glyph? Cappelli shows H as = Roman numeral 200 (p. 214) — any connection???
- Victoriatus: As a coin, this was produced only for a short time, from about 221 BC to 170 BC. I have never seen a special symbol for this coin but will research it further. Fortunately U+0245 LATIN CAPITAL LETTER TURNED V is encoded and has the appropriate appearance. Lewis & Short cites three instances of the term as an apothecary’s weight, including one


example from Marcus Empiricus. I'm guessing that the word continued in use as a weight long after the coins of the same name were no longer produced.

V. Items that require more information


In order to keep everything in one file, I have listed below characters that have come in but about which I do not know enough to make any suggestions or which require confirmation.

The following come from a list sent in by Maurizio in December 2017.

Liber coloniarum

I: actus n. X 

I: aut [] deun. [=XI]

I: , XI] per longum


The first of these looks like the ROMAN NUMERAL SIX LATE FORM, which is encoded at U+2185 C. Does XG mean 'sixteen' in this context? If so, then 2185 is correct.

The others seem like glyph variants of deunx, which is already encoded; need confirmation.

From Marcus Empiricus *De medicamentis*:


VIII 58: 


What do these mean? For the first one, U+A751 p might be appropriate but we need to know the meaning to be sure.


VIII 58: 

XXIV 7: 

From Pelagonius, *Ars veterinaria*

204: aquam sulphuris dragmae 

351 Bo: cerae 

351 Bo: olei 


I consulted the text of Pelagonius. If I am understanding correctly what I found there, the first of these represents *una*, which is probably just the horizontal line sign for an uncia (with its typical value of one). The second represents a semuncia and the third a selibra. Sigma variants are common for the semuncia (see table above). Perhaps the dot in both of these is meant to indicate an abbreviation for a value; if so, we can add glyph variants to the table. Assuming I've got this right and uncia and semuncia are meant, encode them with 10191 and 10192 as usual. The third one is more difficult; please confirm that selibra is the correct reading. We don't have a sign for this.

518: καθαρὸν 

could this be a Greek litra sign, U+10183?? Reference shape in Unicode is ͵

530: οἶνον 

is this the xestes sign but in less good handwriting than the following?

531: πισάνης 

this clearly looks like the Greek xestes sign, U+10185 ϗ; please confirm that this makes sense in the context

© 2020 David J. Perry and Unicode, Inc. All Rights Reserved. The Unicode Consortium and David J. Perry make no expressed or implied warranty of any kind, and assume no liability for errors or omissions. No liability is assumed for incidental and consequential damages in connection with or arising out of the use of the information or programs contained or accompanying this technical note. The Unicode Terms of Use apply.

Unicode and the Unicode Logo are registered trademarks of Unicode, Inc. in the United States and other countries.