



Manuscripts in the Digital Age: XML-Based Catalogues and Editions

European Summer University 2021

Katarzyna Anna Kapitan

@KatarzynaAnn

&

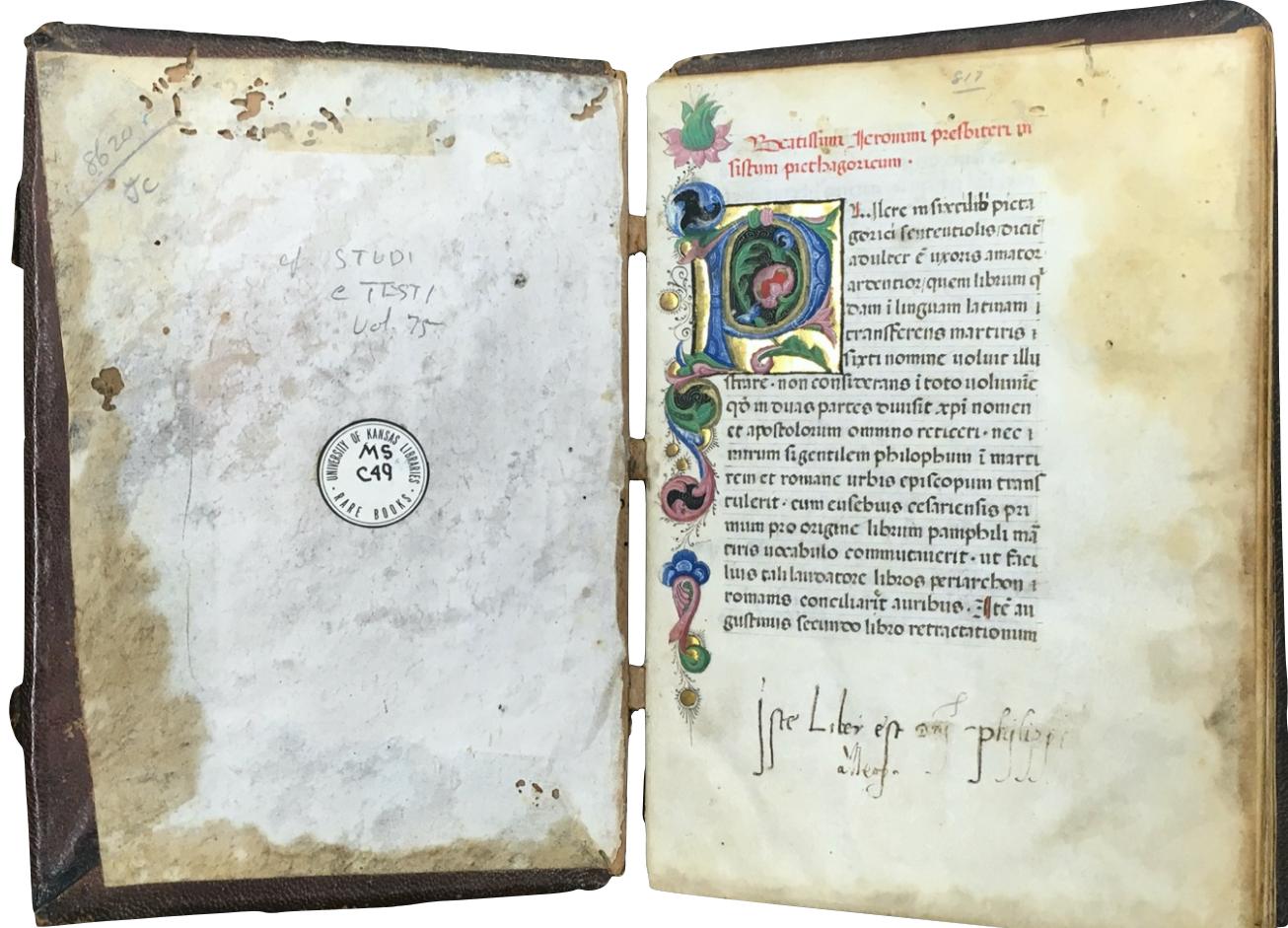
N. Kivilcim Yavuz

@nkivilcimyavuz

Marta Minujín, "The Parthenon of Books." Site-specific installation for Documenta 14, Friedrichsplatz, Kassel, Germany, 2017.

TEI Text Structure Module

HOW TO TRANSCRIBE
THE TEXT OF A MANUSCRIPT
ACCORDING TO
THE TEI GUIDELINES



United States, Lawrence, KS, University of Kansas, Kenneth Spencer Research Library, MS C49, fol. 1r. Photo: N. Kivilcım Yavuz.

P5 Guidelines — English ▾

Search

P5: Guidelines for Electronic Text Encoding and Interchange

Version 4.2.2. Last updated on 9th April 2021, revision 609a109b1

[English] [Deutsch] [Español] [Italiano] [Français] [日本語] [한국어] [中文]



Front Matter

Title

- i. [Releases of the TEI Guidelines](#)
- ii. [Dedication](#)
- iii. [Preface and Acknowledgments](#)
- iv. [About These Guidelines](#)
- v. [A Gentle Introduction to XML](#)
- vi. [Languages and Character Sets](#)

Back Matter

- ⊕ Appendix A [Model Classes](#)
- ⊕ Appendix B [Attribute Classes](#)
- ⊕ Appendix C [Elements](#)
- ⊕ Appendix D [Attributes](#)
- ⊕ Appendix E [Datatypes and Other Macros](#)
- ⊕ Appendix F [Bibliography](#)
- ⊕ Appendix G [Deprecations](#)
- ⊕ Appendix H [Prefatory Notes](#)
- Appendix I [Colophon](#)

Text Body

- ⊕ 1 [The TEI Infrastructure](#)
- ⊕ 2 [The TEI Header](#)
- ⊕ 3 [Elements Available in All TEI Documents](#)
- ⊕ 4 [Default Text Structure](#)
- ⊕ 5 [Characters, Glyphs, and Writing Modes](#)
- ⊕ 6 [Verse](#)
- ⊕ 7 [Performance Texts](#)
- ⊕ 8 [Transcriptions of Speech](#)
- ⊕ 9 [Dictionaries](#)
- ⊕ 10 [Manuscript Description](#)
- ⊕ 11 [Representation of Primary Sources](#)
- ⊕ 12 [Critical Apparatus](#)
- ⊕ 13 [Names, Dates, People, and Places](#)
- ⊕ 14 [Tables, Formulæ, Graphics and Notated Music](#)
- ⊕ 15 [Language Corpora](#)
- ⊕ 16 [Linking, Segmentation, and Alignment](#)
- ⊕ 17 [Simple Analytic Mechanisms](#)
- ⊕ 18 [Feature Structures](#)
- ⊕ 19 [Graphs, Networks, and Trees](#)
- ⊕ 20 [Non-hierarchical Structures](#)
- ⊕ 21 [Certainty, Precision, and Responsibility](#)
- ⊕ 22 [Documentation Elements](#)
- ⊕ 23 [Using the TEI](#)

TEI sourcecode

- [Getting and Using the TEI Sources.](#)
- [TEI GitHub Repository](#)
- [Bug Reports, Feature Requests, etc.](#)

Text Structure

<text> (text) contains a single text of any kind, whether unitary or composite, for example a poem or drama, a collection of essays, a novel, a dictionary, or a corpus sample.

<front> (front matter) contains any prefatory matter (headers, abstracts, title page, prefaces, dedications, etc.) found at the start of a document, before the main body.

<body> (text body) contains the whole body of a single unitary text, excluding any front or back matter.

<group> contains the body of a composite text, grouping together a sequence of distinct texts (or groups of such texts) which are regarded as a unit for some purpose, for example because they are part of a composite manuscript.

<back> (back matter) contains any appendixes, etc. following the main part of a text.

<text>: How it works

```
<teiHeader>
  <sourceDesc>
    <msDesc>
      <msIdentifier> </msIdentifier>
      <msContents> </msContents>
      <physDesc> </physDesc>
      <history> </history>
      <additional> </additional>
    </msDesc>
  </sourceDesc>
</teiHeader>
<text>
  <front> </front>
  <body> </body>
  <back> </back>
</text>
```

<text>: How it works

```
<text>
  <front> </front>
  <group>
    <text>
      <front> </front>
      <body> </body>
      <back> </back>
    </text>
    <text>
      <front> </front>
      <body> </body>
      <back> </back>
    </text>
  </group>
  <back> </back>
</text>
```

<front><body><back>

<**div**> (text division) contains a subdivision of the front, body, or back of a text.

@type characterizes the element in some sense, using any convenient classification scheme or typology.

@subtype provides a sub-categorization of the element, if needed.

@n gives a number (or other label) for an element, which is not necessarily unique within the document.

Fundamental Elements of Prose and Verse

< p > (paragraph) marks paragraphs in prose.

< lg > (line group) contains one or more verse lines functioning as a formal unit, e.g. a stanza, refrain, verse paragraph, etc.

< l > (verse line) contains a single, possibly incomplete, line of verse.

@met (metrical structure, conventional) contains a user-specified encoding for the conventional metrical structure of the element.

@rhyme (rhyme scheme) specifies the rhyme scheme applicable to a group of verse lines.

Layout: Milestone Elements

<**gb**> (gathering beginning) marks the beginning of a new gathering or quire in a transcribed codex.

<**pb**> (page beginning) marks the beginning of a new page in a paginated document.

<**cb**> (column beginning) marks the beginning of a new column of a text on a multi-column page.

<**lb**> (line beginning) marks the beginning of a new (typographic) line in some edition or version of a text.

Additions, Deletions, and Omissions

<add> (addition) contains letters, words, or phrases inserted in the source text by an author, scribe, or a previous annotator or corrector.

**** (deletion) contains a letter, word, or passage deleted, marked as deleted, or otherwise indicated as superfluous or spurious in the copy text by an author, scribe, or a previous annotator or corrector.

<gap> indicates a point where material has been omitted in a transcription because the material is illegible, invisible, or inaudible.

<unclear> contains a word, phrase, or passage which cannot be transcribed with certainty because it is illegible or inaudible in the source.

<supplied> signifies text supplied by the transcriber or editor for any reason; for example because the original cannot be read due to physical damage, or because of an obvious omission by the author or scribe.

Apparent Errors

<**sic**> (Latin for thus or so) contains text reproduced although apparently incorrect or inaccurate.

<**corr**> (correction) contains the correct form of a passage apparently erroneous in the copy text.

Regularization and Normalization

<**reg**> (regularization) contains a reading which has been regularized or normalized in some sense.

<**orig**> (original form) contains a reading which is marked as following the original, rather than being normalized or corrected.

Abbreviations

<**abbr**> (abbreviation) contains an abbreviation of any sort.

<**expan**> (expansion) contains the expansion of an abbreviation.

<**am**> (abbreviation marker) contains a sequence of letters or signs present in an abbreviation which are omitted or replaced in the expanded form of the abbreviation.

<**ex**> (editorial expansion) contains a sequence of letters added by an editor or transcriber when expanding an abbreviation.

<abbr>: Example

<p>How are you doing <abbr>Dr.</abbr> Jones?</p>

<p>How are you doing <abbr>Dr<am>. </am></abbr> Jones?</p>

<p>How are you doing <abbr>Dr<am>. </am></abbr>

<expan>D<ex>octo</ex>r</expan>Jones?</p>

<choice>

<**sic**> (Latin for thus or so) contains text reproduced although apparently incorrect or inaccurate.

<**corr**> (correction) contains the correct form of a passage apparently erroneous in the copy text.

<**reg**> (regularization) contains a reading which has been regularized or normalized in some sense.

<**orig**> (original form) contains a reading which is marked as following the original, rather than being normalized or corrected.

<**unclear**> contains a word, phrase, or passage which cannot be transcribed with certainty because it is illegible or inaudible in the source.

<**supplied**> signifies text supplied by the transcriber or editor for any reason; for example because the original cannot be read due to physical damage, or because of an obvious omission by the author or scribe.

<choice>: Example

```
<choice>
    <sic> the text as it is on the
    manuscript </sic>
    <corr> the corrected text </corr>
</choice>
```

<choice>

<**abbr**> (abbreviation) contains an abbreviation of any sort.

<**expan**> (expansion) contains the expansion of an abbreviation.

<**am**> (abbreviation marker) contains a sequence of letters or signs present in an abbreviation which are omitted or replaced in the expanded form of the abbreviation.

<**ex**> (editorial expansion) contains a sequence of letters added by an editor or transcriber when expanding an abbreviation.

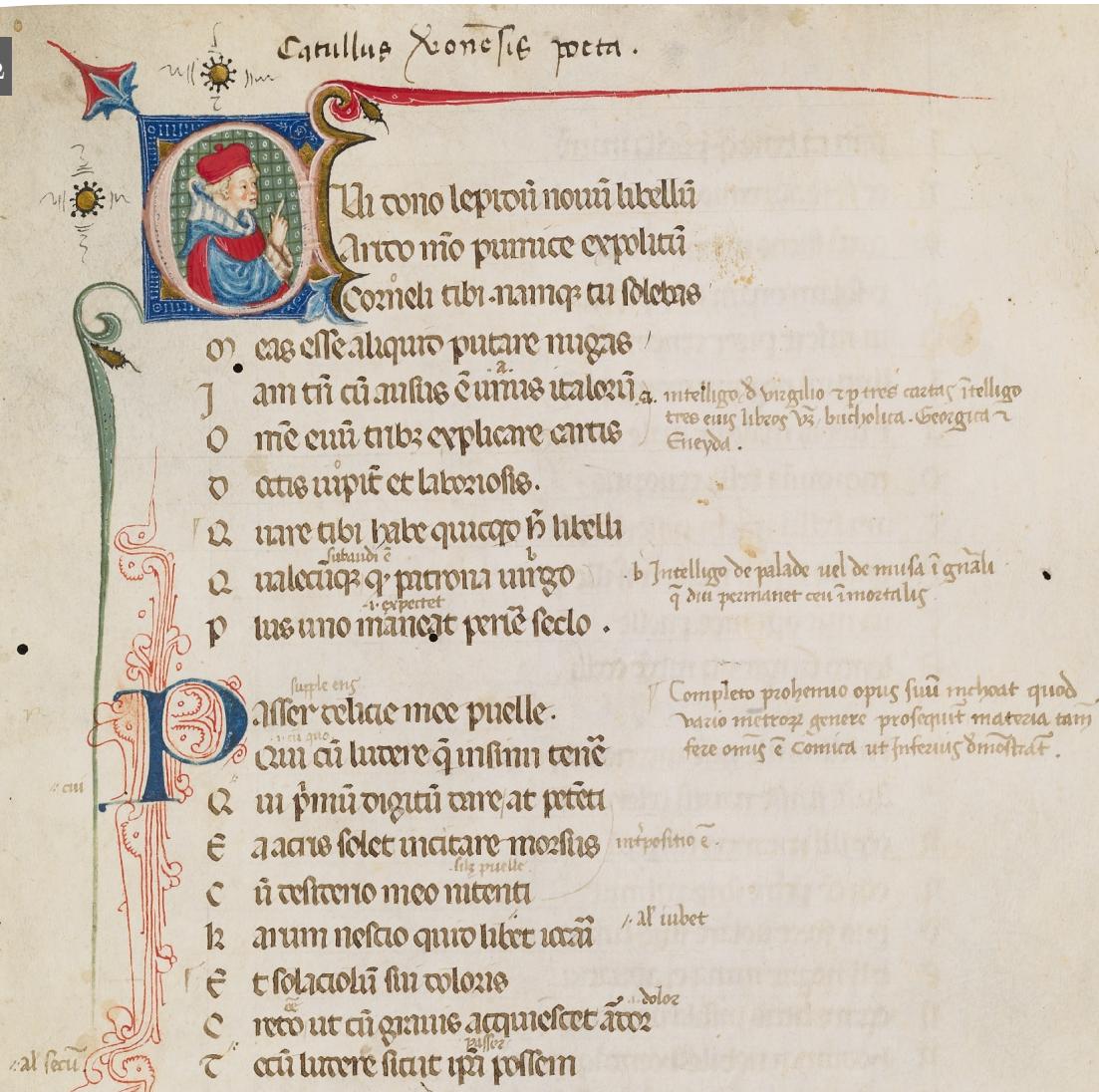
Transcription: Use of attributes

@reason gives the reason for omission or indicates why the material is hard to transcribe. Value examples: cancelled, illegible, faded, etc.

Can be used as part of <gap> <unclear> <supplied>

<text> and Text Structure





United Kingdom, Oxford, Bodleian Library, Canon. Class. Lat. 30, 1r
(detail).

Encoding Texts in Verse According to the TEI Guidelines

POEMS NO. 1 & 2 BY CATULLUS

HANDS ON





1. Open Oxygen XML Editor.
2. Download and open the files from the Google Drive folder

Week 2 > Day 1 > Exercises:

- MUFI Entities for Hands-On Encoding Texts in Verse
- “starter” xml files and images for three manuscripts

Catullus Online

CATULLUS ONLINE

AN ONLINE REPERTORY OF CONJECTURES ON CATULLUS



[!\[\]\(e7cf99a695198f48d7c53d4da8381241_img.jpg\) facebook](#)

POEMS

POEMS WITH APPARATUS

BIBLIOGRAPHY

TESTIMONIA

SEARCH:

search in poems 

search in apparatus

search in testimonia

GO TO PASSAGE:

poem - line 

MANUSCRIPTS

ABOUT THE WEBSITE

ABOUT THE REPERTORY

About the website

This website offers a critical edition of the poems of Catullus, a repertory of conjectures on the text, an overview of the ancient quotations from Catullus that have independent source value, and high-quality images of some of the most important manuscripts. It was constructed between 2009 and 2013 in the course of the research project *An Online Repertory of Conjectures for Catullus* at the [Center for Advanced Studies](#) and the [Abteilung für Griechische und Lateinische Philologie](#) of the [Ludwig-Maximilians-Universität München](#). It was conceived and designed, and its contents were assembled by [Dániel Kiss](#), who continues to act as its editor. The technical framework of the site was built by [Woodpecker Software](#), while [Stalker Studio](#) were responsible for the graphic design.

Most of the research that has led to the repertory was conducted at the branches of the [Universitätsbibliothek of the Ludwig-Maximilians-Universität München](#), and at the [Bayerische Staatsbibliothek](#). Further research was done in Munich at the library of [Monumenta Germaniae Historica](#), that of the [Thesaurus Linguae Latinae](#), and that of the [Franziskanerkloster St. Anna im Lehel](#); and at the [Staats- und Stadtbibliothek](#) in Augsburg; the [Harry Ransom Center](#) in Austin (Texas); the library of the [Ibero-Amerikanisches Institut-Preussischer Kulturbesitz](#), the [Staatsbibliothek zu Berlin-Preussischer Kulturbesitz](#), and the [Universitätsbibliothek](#) of the [Humboldt-Universität](#) in Berlin; the [Koninklijke Bibliotheek-Bibliothèque Royale](#) in Brussels; the [Országos Széchényi Könyvtár](#) in Budapest; the [Davis Library](#), the [Ullman Library](#), and the [Wilson Library](#) of the University of North Carolina at Chapel Hill; the [Biblioteka Jagiellońska](#) in Cracow; the [Biblioteca Medicea](#), [Biblioteca Nazionale Centrale](#), and the [Biblioteca Riccardiana](#) in Florence;

<http://www.catullusonline.org> by Dániel Kiss

Edition by E. T. Merrill on Perseus



C. Valerius Catullus, *Carmina*
E. T. Merrill, Ed.

Search
("Agamemnon", "Hom. Od. 9.1", "denarius")
[All Search Options](#) [view abbreviations]

[Home](#) [Collections/Texts](#) [Perseus Catalog](#) [Research](#) [Grants](#) [Open Source](#) [About](#) [Help](#)

Your current position in the text is marked in blue. Click anywhere in the line to jump to another position. [Hide browse bar](#)

poem:



This text is part of:
[Greek and Roman Materials](#)
[Latin Poetry](#)
[Latin Texts](#)
[Catullus](#)

Search the Perseus Catalog for:
[Editions/Translations](#)
[Author Group](#)

View text chunked by:
[poem](#) : [line](#)

Table of Contents:
[poem 1](#)
[poem 2](#)
[poem 3](#)
[poem 4](#)



Click on a word to bring up parses, dictionary entries, and frequency statistics

Catul. 1

Cui dono lepidum novum libellum
arido modo pumice expolitum?
Corneli, tibi; namque tu solebas
meas esse aliquid putare nugas,
iam tum cum ausus es unus Italorum
omne aevum tribus explicare chartis,
doctis, Iuppiter, et laboriosis!
quare habe tibi quidquid hoc libelli
qualecumque, quod, o patrona virgo,
plus uno maneat perenne saeclo.

The National Endowment for the Humanities provided support for entering this text.

Purchase a copy of this text (not necessarily the same edition) from [Amazon.com](#)



This work is licensed under a [Creative Commons Attribution-ShareAlike 3.0 United States License](#).

An [XML version](#) of this text is available for download, with the additional restriction that you offer Perseus any modifications you make. Perseus provides credit for all accepted changes, storing new additions in a versioning

Notes (E. T. Merrill, 1893) [focus load](#)

English (Sir Richard Francis Burton, 1894) [focus load](#)

English (Leonard C. Smithers, 1894) [focus load](#)

References (13 total) [hide](#)

- Commentary references to this page (8):
 - E. T. Merrill, *Commentary on Catullus*, 100
 - E. T. Merrill, *Commentary on Catullus*, 102
 - E. T. Merrill, *Commentary on Catullus*, 14b
 - E. T. Merrill, *Commentary on Catullus*, 23
 - E. T. Merrill, *Commentary on Catullus*, 40
 - E. T. Merrill, *Commentary on Catullus*, 49
 - E. T. Merrill, *Commentary on Catullus*, 6
 - E. T. Merrill, *Commentary on Catullus*, 78
- Cross-references to this page (5):
 - E. T. Merrill, *Commentary on Catullus*, Friends and foes.
 - E. T. Merrill, *Commentary on Catullus*, Metres.
 - E. T. Merrill, *Commentary on Catullus*, Education
 - Anne Mahoney, *Overview of Latin Syntax, Nouns, Adjectives, and Pronouns*
 - Anne Mahoney, *Overview of Latin Syntax, Verbs*

Vocabulary Tool [load](#)

Search [hide](#)

Texts in Verse: Basic Structure

```
<text>
  <body>
    <div>
      <p>
        <lg>
          <l> </l>
          <l> </l>
          <l> </l>
          <l> </l>
        </lg>
      </p>
    </div>
  </body>
</text>
```

UNICODE



">\$	¤	♦	◆	😊	🎲	∞	}`	□	♪	😢	
U+00A2	U+05E1	U+2666	U+1F617	U+1F3B2	U+221E	U+2307	U+2752	U+266B	U+2639		
‘	˶	♥	>	🍾	😱	⸮	😿	§	﴿		
U+2018	U+1F47D	U+1F49C	U+203A	U+1F37E	U+1F647	U+0A09	U+1F639	U+00A7	U+1886		
Adopt a Character	+	▫	♡	Everyone in the world should be able to use their own language on phones and computers.							
Emoji	+	⊗	϶	LEARN MORE ABOUT UNICODE						ADOPT A CHARACTER ↗	
Basic Info	+	♀	϶								
Latest	+	😊	✗	😺	❓	→	⚡	✌	🧙	⌚	
Connect	+	U+1F604	U+00D7	U+1F431	U+203D	U+2192	U+1F923	U+270C	U+261E	U+1F9D9	U+00A9
Membership	+	★	♥	፩	😭	☺	☎	🧩	﴿	♪	█
Press		U+2605	U+1F9E1	U+0937	U+1F625	U+1F609	U+260E	U+1F9E9	U+2318	U+266A	U+2B1B
		U+1F4A2	U+1F600	U+2602	≈	😊	🎾	Ψ	♪	💯	滿
							U+03A8	U+266C	U+1F4AF		U+1F235




[Home](#)
[Browse by character](#)
[Browse by code chart](#)
[Browse by range](#)
[Browse by updates](#)
[Full code chart](#)
[MUFI: The Medieval Unicode Font Initiative](#)


MUFI: The Medieval Unicode Font Initiative

Archived old MUFI site

Search the character database using the icon at the top right, or browse using the links on the left (or icon at the top left on smaller devices).

Download a snapshot of MUFI's data as: [csv](#) or [json](#) (CC BY-SA 4.0).

Disclaimer

This site is managed by scholars in Medieval studies with the aim of establishing a consensus on the use of Unicode among medievalists. It is not affiliated with or endorsed by Unicode.

Board 2016-

subsections

1. Background



2. MUFI character recommendation



3. MUFI fonts



4. Medievalist proposals to Unicode



5. LINCUA



6. Board 2001–2015



7. Other Fonts & Projects



8. Links

<https://mufi.info/>

Texts in Verse: Abbreviations

MUFI Entity	Character	Name
&bar;	—	Combining abbreviation mark bar above
&barbl;	—	Combining abbreviation mark bar below
&etfin;	ȝ	Latin abbreviation sign small final et (3-shaped mark)
&er;	᳚	Combining abbreviation mark superscript er
&ercurl;	՚	Combining abbreviation mark zigzag above curly
&qbardes;	᳜	Latin small letter q with stroke through descender
&rsup;	᳝	Combining latin small letter r
ı	᳞	Latin small letter dotless i
&inodotsup;	᳟	Combining Latin small letter dotless i
&rum;	᳡	Latin abbreviation sign small rum
&	᳢	Ampersand
&et;	᳣	Tironian sign et

Poems no. 1 & 2 by Catullus

Group 1: Dresden, Sächsische Landesbibliothek – Staats- und Universitätsbibliothek, Mscr.Dresd.Dc.133

Group 2: Oxford, Bodleian Library, Canon. Class. Lat. 30

Group 3: Paris, Bibliothèque nationale de France, latin 14137