

Data Entry Specs 1.1.2 Overview

Wolfgang Schmidle, Klaus Thoden, Malcolm D. Hyman
Max Planck Institute for the History of Science, Berlin, Germany
24th October 2008

2 General Markup

2.1 Pages

Type the entire contents of one page, then go on to the next page. Do not mix the contents of different pages.

2.1.1 Page Breaks, Page Numbers and Running Heads

Page breaks are marked by `<pb>`. If the page has a page number, type it within the `<pb>` tag, e.g. `<pb 6>`. Type the page number exactly as it appears in the book. If there is a running head on the page, it is marked by `<rh>` and `</rh>`. Type the running head immediately after the `<pb>` tag.

2.1.2 Catchwords and Signatures

Do not type catchwords or signatures.

2.2 Text Blocks

Type a return after each line of the printed page.

2.2.1 Headings

Headings are marked by `<h>` and `</h>`.

2.2.2 Paragraphs

Paragraphs are marked by `<p>` and `</p>`.

2.2.3 Block Quotations

A block quotation is marked by `<q>` and `</q>`. Do not type repeating quotation symbols.

2.3 Columns and Tables

2.3.1 Columns

Columns are marked by `<col>` and `</col>`. Assign a number to each column and type it in the `<col>` tag.

2.3.2 Tables

A table is marked by `<tb>` and `</tb>`. Use # as field separators. Type a return after each row. Do not type horizontal or vertical lines.

2.4 Notes

2.4.1 Marginal Notes

A note in the left margin is marked by `<mg1>` `</mg1>`, and a note in the right margin is marked by `<mgr>` `</mgr>`. Type the marginal note on separate lines, starting after the line it is closest to.

2.4.2 Footnotes

Footnotes are marked by `<fn>` and `</fn>`. Type the footnote where it appears on the page. Insert the footnote symbol or number in the `<fn>` tag. The corresponding footnote symbol in the main text is marked by `<n>`. Insert the footnote symbol or number in the `<n>` tag.

2.5 Printed Images

2.5.1 Figures

Where a figure occurs in the text, type a `<fig>` tag on a separate line. If you can identify a caption of the figure, mark it by `<cap>` `</cap>`.

2.5.2 Drop Caps

Type a drop cap as an ordinary capital letter.

2.6 Handwritten Notes

Handwritten notes are marked by `<hd>`. Do not type the note itself.

2.7 Unreadable Text

2.7.1 Characters You are Unsure About

If you are not sure about a character, type `<?>` after it. If you are unsure about a whole paragraph, type `<?>` directly after the `<p>` tag, i.e. `<p><?>`. A completely unreadable character is typed as `@`. If it is unclear how many characters are unreadable, use `<gap>` instead of `@`.

2.7.2 Unknown Characters

If there is an unknown character in the text, add it to the numbered list of unknown characters. From then on, type its number whenever it occurs in the text, e.g. `<001>`.

3 Latin Alphabet

3.1 General

3.1.1 Punctuation

Type a space after the punctuation marks `. , : ; ! ?` Type a space before the opening brackets `([` and after the closing brackets `)]`

3.1.2 Hyphens

If the line ends with a hyphen, type it.

3.2 Characters

3.2.1 Characters to be Typed Directly

If a character with diacritics can be easily typed and encoded as a single Unicode character, type it directly.

3.2.2 Other Diacritics

If a character cannot be typed directly but is a combination of a known character and a known diacritic, type `\`, then the diacritic, then the character.

3.3 Type Styles

3.3.1 Italics

Use underscores `_ _` to mark the beginning and end of words or whole lines in italics. Encode only up to a few lines of text this way. A whole paragraph in italics is marked by `it` in the `<p>` tag, i.e. `<p it>`.

3.3.3 Small Caps

If you can identify capital letters as small caps, use `<sc>` `</sc>`.

3.3.4 Subscript and Superscript

Use `<_>` and `</_>` for subscript. Use `<^>` and `</^>` for superscript.

3.4 Latin Ligatures

Resolve easy ligatures silently. Resolve difficult ligatures too, but type `{ and }` around them.

4 Greek Alphabet

4.1 Characters

Type Greek characters directly as Unicode characters.

4.2 Punctuation

The rules for Latin punctuation apply. In addition, type the mid-dot `·` directly.

4.3 Greek Ligatures

Resolve letter variations silently. Resolve all ligatures and type `{ and }` around them. If a ligature contains some diacritics, type them.

5 Mathematics

5.1 Mathematical Symbols

Type common mathematical symbol directly as Unicode characters.

5.2 Fractions

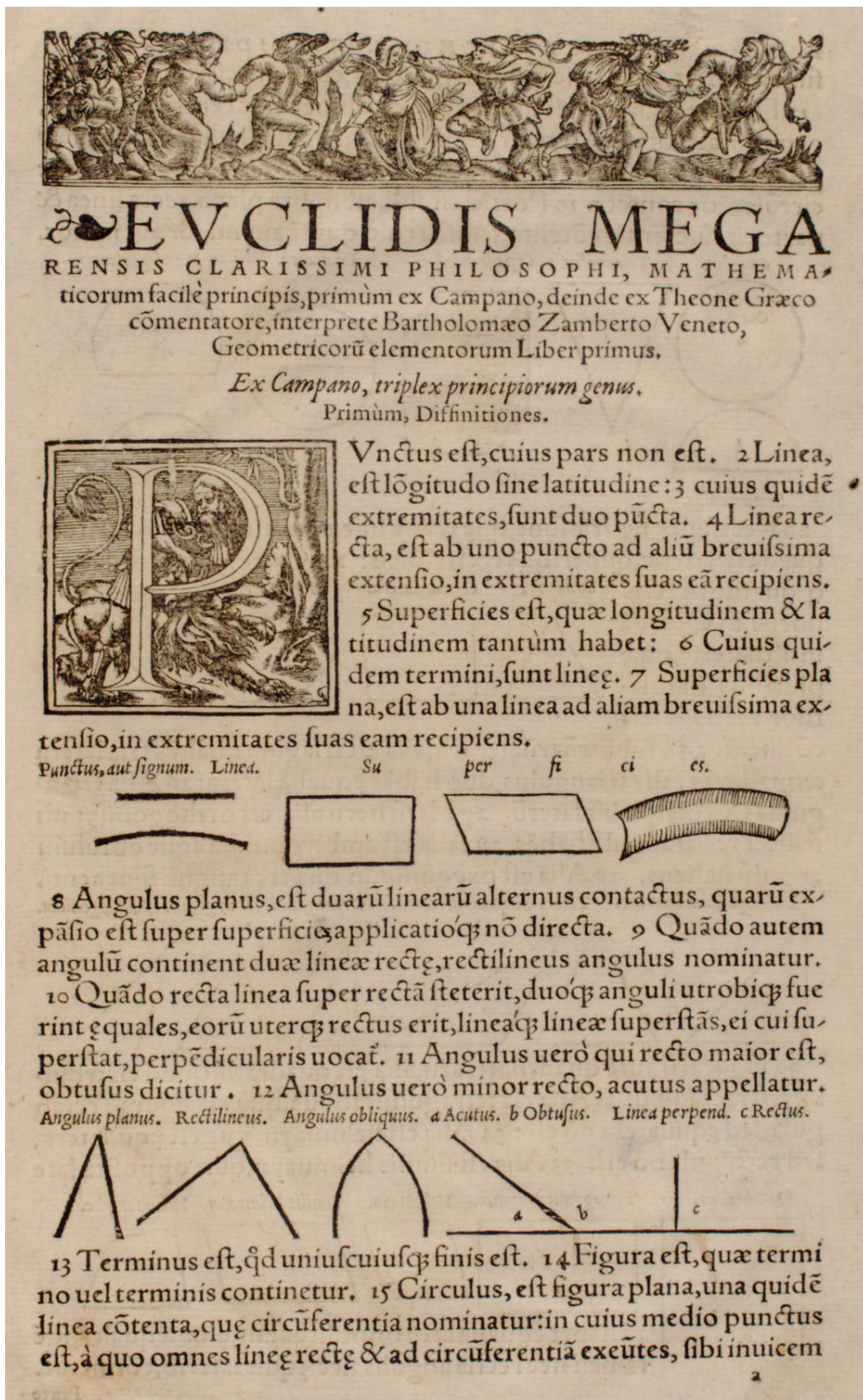
Type fractions in one line. Use { / } to mark beginning, fraction line and ending.

A List of All Tags

section	tag	name	may contain
2.1.1	<pb>	page break	page number, it
2.1.1	<rh> </rh>	running head	it
2.2.1	<h> </h>	heading (or footer)	it
2.2.2	<p> </p>	paragraph	it
2.2.3	<q> </q>	block quotation	it
2.3.1	<col> </col>	column	column number, it
2.3.2	<tb> </tb>	table	it
2.3.2	#	field separator	
2.4.1	<mgl> </mgl>	marginal note (left)	anchor symbol
2.4.1	<mgr> </mgr>	marginal note (right)	anchor symbol
2.4.2	<n>	footnote (main text)	footnote symbol
2.4.2	<fn> </fn>	footnote	footnote symbol
2.5.1	<fig>	figure	
2.5.1	<cap> </cap>	figure caption	it
2.6	<hd>	handwritten note	
2.7.1	@, <gap>	unreadable text	
2.7.1	<?>	uncertain text	
2.7.2	<001>, etc.	unknown character	
3.2.2	\' q, etc.	character+diacritic	' \ ^ " ~ , . = -
3.3.1	<u> </u>	italics	
3.3.2	<bf> </bf>	bold face	
3.3.3	<sc> </sc>	small caps	
3.3.4	<_> </_>	subscript	
3.3.4	<^> </^>	superscript	
3.3.5	<red> </red>	text in red	
3.4	{quo}d, etc.	resolved ligature	
5.2	{ / }	fraction	

B Examples

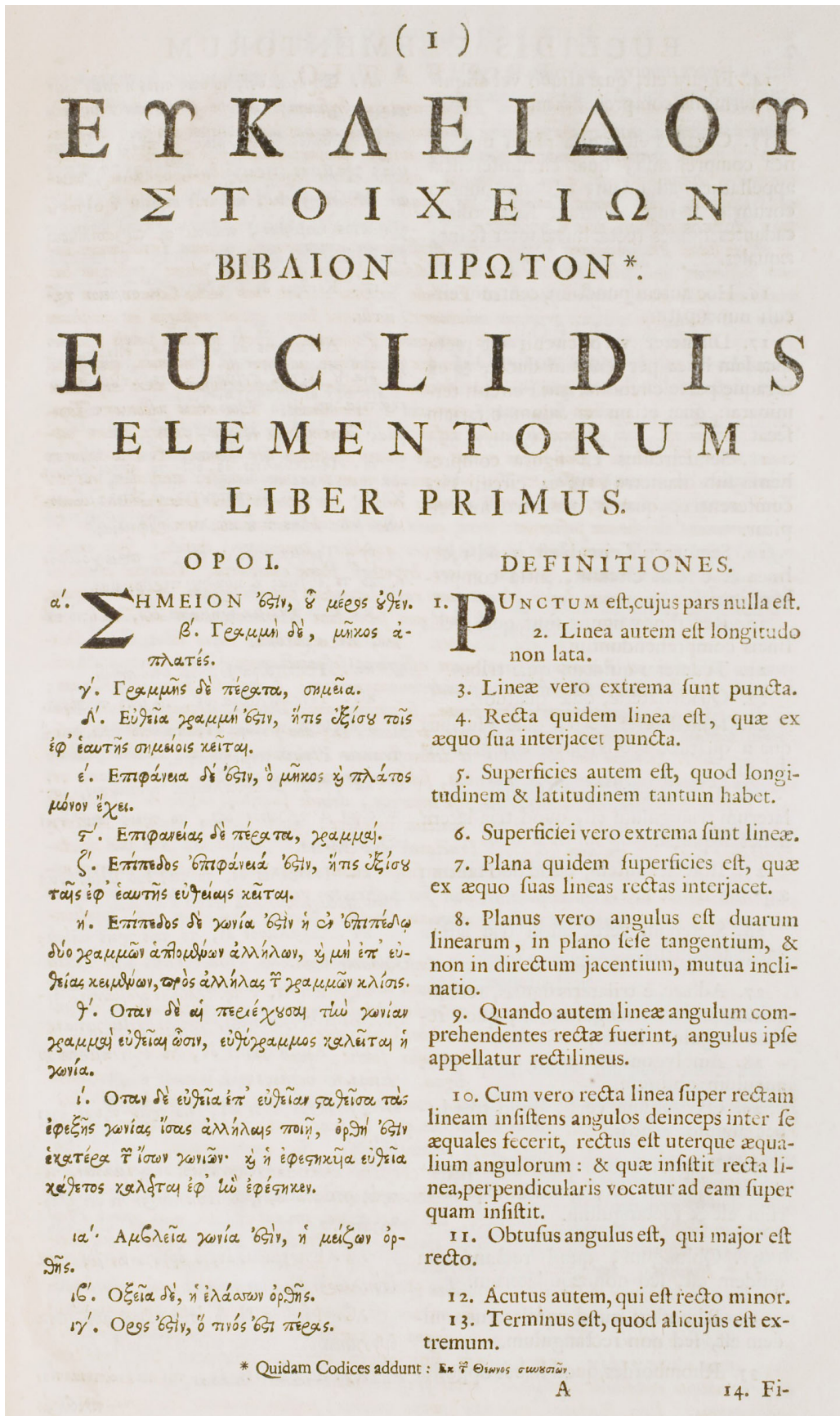
B.1 Latin Example



<pb>
 <fig>
 <fig>
 <h>EVCLIDIS MEGA
 RENSIS CLARISSIMI PHILOSOPHI, MATHEMA-
 ticorum facilè principis, primùm ex Campano, deinde ex Theone Græco
 cõmentatore, interprete Bartholomæo Zamberto Veneto,
 Geometricorũ elementorum Liber primus.</h>
 <h it>Ex Campano, triplex principiorum genus.</h>
 <h>Primùm, Diffinitiones.</h>
 <p>PVnctus eſt, cuius pars non eſt. 2 Linea,
 eſt lōgitudō ſine latitudine: 3 cuius quidē
 extremitates, ſunt duo p~ucta. 4 Linea re-
 cta, eſt ab uno puncto ad ali~u breuiſſima
 extenſio, in extremitates ſuas eã recipiens.</p>
 <p>5 Superficies eſt, quæ longitudinem & la-
 titudinem tantùm habet: 6 Cuius qui-
 dem termini, ſunt line{ae}. 7 Superficies pla-
 na, eſt ab una linea ad aliam breuiſſima ex-
 tenſio, in extremitates ſuas eam recipiens.</p>
 <cap it>Punctus, aut ſignum. Linea.</cap>
 <fig>
 <cap it>Su per fi ci es.</cap>
 <fig><fig><fig>
 <p>8 Angulus planus, eſt duar~u linear~u alternus contactus, quar~u ex-
 pãſio eſt ſuper ſuperficies,<?> applicatio´{que} nõ directa. 9 Quãdo autem
 angul~u continent duæ lineæ rect{ae}, rectilineus angulus nominatur.</p>
 <p>10 Quãdo recta linea ſuper rectã ſteterit, duo´{que} anguli utrobi{que} fue-
 rint {ae}quales, eor~u uter{que} rectus erit, linea´{que} lineæ ſuperſtãs, ei cui ſu-
 perſtat, perp~edicularis uoca~t. 11 Angulus uerò qui recto maior eſt,
 obtuſus dicitur. 12 Angulus uerò minor recto, acutus appellatur.</p>
 <cap it>Angulus planus.</cap>
 <fig>
 <cap it>Rectilineus.</cap>
 <fig>
 <cap it>Angulus obliquus.</cap>
 <fig>
 <cap it>a Acutus. b Obtuſus. Linea perpend. c Rectus.</cap>
 <fig>
 <p>13 Terminus eſt, {quo}d uniuſcuius{que} finis eſt. 14 Figura eſt, quæ termi-
 no uel terminis continetur. 15 Circulus, eſt figura plana, una quid~e
 linea cõtenta, qu{ae} circ~uferentia nominatur: in cuius medio punctus
 eſt, à quo omnes line{ae} rect{ae} & ad circ~uferentiã exe~utes, ſibi inuicem

PLEASE NOTE: The typesetter used the page number of the next page as a catchword, i.e. it is not the page number of this page. The closing </p> of the last paragraph is on the next page.

B.2 Greek Example



<pb (1)>
 <h>ΕΥΚΛΕΙΔΟΥ
 ΣΤΟΙΧΕΙΩΝ
 ΒΙΒΛΙΟΝ ΠΡΩΤΟΝ <n *>.</h>
 <h>EUCLIDIS
 ELEMENTORUM
 LIBER PRIMUS.</h>
 <col 1>
 <h>ΟΡΟΙ.</h>
 <p>α'. ΣΗΜΕΙΟΝ {έστι}ν, {οῦ} μέ{ρο}ς {οῦ}{θέ}ν.</p>
 <p>β'. Γ{ρα}μ{μῆ} {δὲ}, {μῆ}{κο}ς ἀ-
 {πλ}ατέ{ς}.</p>
 <p>γ'. Γ{ρα}μ{μῆ}ς {δὲ} {πέ}{ρα}{τα}, {ση}{μ{εῖ}α}.</p>
 <p>δ'. Εὐ{θε}ῖα {γρ}αμ{μῆ} {έστι}ν, ἣ{τι}ς {έξ}ί{σ{ου} {το}ίς
 έφ' έ{αυ}τῆς {ση}{μ{εῖ}οις {κε}ῖ{ται}.</p>
 <p>ε'. Ε{πι}φάν{ει}α {δέ} {έστι}ν, ὁ {μῆ}{κο}ς {καί} {πλ}ά{το}ς
 {μό}νον ἔχ{ει}.</p>
 <p>ς'. Ε{πι}φ{αν}{εῖ}{α}ς {δὲ} {πέ}{ρα}{τα}, {γρ}αμ{αί}.</p>
 <p>ζ'. Ε{πί}{πε}{δο}ς {έπι}φάν{ει}ά {έστι}ν, ἣ{τι}ς {έξ}ί{σ{ου} {τ{αῖ}ς
 έφ' έ{αυ}τῆς εὐθ{εῖ}{α}ις κ{εῖ}τ{αί}.</p>
 <p>η'. Ε{πί}{πε}{δο}ς {δὲ} {γω}νία {έστι}ν ἢ {έν} {έπι}{πέδ}ω
 {δύ}ο {γρ}αμ{μῶ}ν ἀ{πτ}ο{μέν}ων ἀ{λλ}ήλων, {καί} {μῆ} έπ' εὐ-
 {θε}ί{α}ς κ{εῖ}{μέν}ων, {πρ}ὸς ἀ{λλ}ήλ{α}ς {τῶν} {γρ}αμμῶν κλί{σι}ς.</p>
 <p>θ'. Ο{τα}ν {δὲ} {αί} {πε}{ρι}έχ{ου}{σαι} {τῆν} {γω}νί{αν}
 {γρ}αμ{μαί} εὐθ{ε}ῖ{α}ι ὄ{σι}ν, εὐθ{ύ}{γρ}αμ{μο}ς {κα}λ{εῖ}τ{αί} ἢ
 {γω}νία.</p>
 <p>ι'. Ο{τα}ν {δὲ} εὐθ{ε}ῖα έπ' εὐθ{ε}ῖ{αν} {στα}{θε}ῖ{σα} {τὰ}ς
 έφεξῆς {γω}νί{α}ς ἴ{σα}ς ἀ{λλ}ήλ{α}ις {πο}ιῆ, ὀρ{θή} {έστι}ν
 έκ{α}τέ{ρα} {τῶν} ἴ{σω}ν {γω}νιῶν· {καί} ἢ έφε{στη}κ{υῖ}α εὐθ{ε}ῖα
 {κά}{θε}ί{το}ς {κα}λ{εῖ}τ{αί} έφ' {ἦν} έφέ{στη}{κε}ν.</p>
 <p>ια'. Αμβλεῖα {γω}νία {έστι}ν, ἢ μ{εῖ}ζων ὀρ-
 {θῆ}ς.</p>
 <p>ιβ'. Οξεῖα {δὲ}, ἢ έλά{σσω}ν ὀρ{θῆ}ς.</p>
 <p>ιγ'. Ο{ρο}ς {έστι}ν, ὁ {τι}νός {έστι} {πέ}{ρα}ς.</p>
 </col 1>
 <col 2>
 <h>DEFINITIONES.</h>
 <p>1. PU<sc>NCTUM</sc> eſt, cujus pars nulla eſt.</p>
 <p>2. Linea autem eſt longitudo
 non lata.</p>
 <p>3. Lineæ vero extrema ſunt puncta.</p>
 <p>4. Recta quidem linea eſt, quæ ex
 æquo ſua interjacet puncta.</p>
 <p>5. Superficiæ autem eſt, quod longi-
 tudinem & latitudinem tantum habet.</p>
 <p>6. Superficiæi vero extrema ſunt lineæ.</p>
 (some untranscribed text)
 <p>12. Acutus autem, qui eſt recto minor.</p>
 <p>13. Terminus eſt, quod alicujus eſt ex-
 tremum.</p>
 </col 1>
 <fn *>Quidam Codices addunt: Εκ {τῶν} Θέωνος σ{υν}{ου}{σι}ῶν.</fn>

PLEASE NOTE: The character ς, which occurs only once on this page (in <p>ς') and stands for the number 6, is not an end-sigma ς, but a stigma. It should be typed as U+03DB (Greek small letter stigma).