

C List of All Trial-and-error Issues

- We will see what they will make out of marginal notes that consist of more than one paragraph.
- Can they identify footnote symbols and the anchors of anchored marginal notes?
- Does `_ _` for single non-italic words in `<p it>`-paragraphs work?
- A page with two columns, and when one column is finished, the other columns uses the space of the other column as well.
- Paragraphs that begin on one page and end on the next page
- `C<sc>AMPANUS<sc>`: Will they identify small caps? Normally they are supposed to ignore different font sizes.

D Design Decisions

Leave the images in colour, but it should also work in b&w.

bold face for tags: the rule is “everything that is not on the page will be printed in bold face”

E Open Questions

Ask the company which characters they can type directly (probably at least the characters from the list in section 3.2.1). For example `ẽ` ?

Most likely, the typists will not only create a list of unknown characters, but also a (written?) list of ad hoc typing conventions where our rules are fuzzy. Can we ask them to send us this list as well?

Structure

columns: I have left out “left to right” in anticipation of Chinese and Arabic. Does that make sense?

How is vertical text encoded? (Georg Agricola, De Re Metallica, p.128)

How is text in red encoded?

Do we lose structural information if drop caps are typed as normal characters? Sometimes they are used as an additional structuring element below the section level (only in modern texts?)

`<n>` tag for footnote symbols in the main text: insert space before and after if there are spaces in the text, i.e treat footnote symbols as punctuation marks? No explicit rule, but how do we do it in our examples? If we use the modern way,¹ we contradict our punctuation rule. → In our examples we insert spaces exactly as in the original text, and until now it works just fine.

Example for a paper crease: remove the rectangle in the image, as it may make other letters more difficult to read?

(Add 3mm before the Example macros, then go through all instances of `\vspace{negative value}`.)

Latin Alphabet

Single and double quotation marks: We only say (in the Block Quote section): “For inline quotations within a paragraph, type the quotation marks exactly as they appear in the text.” Does that mean only the position, or are they indeed supposed to type the quotation mark exactly as it looks like? As Unicode characters? (Do we need a section about quotation marks?)

example in paragraph section (p.5): Is it “feipfo” or “fe ipfo”?

Replace the punctuation example (a weird table, in text form, of Pythagorean triples)?

Table of characters to be typed directly: add more characters? → Probably we will stick to this core list.

Other diacritics: We make them type acute, grave, circumflex, umlaut/diaeresis, tilde and cedilla with backslash escape sequences. Ogonek and ring above: one character each is in the ligature list, and so far there are no additional examples. I have added an escape sequence for “dot above” (section 3.2.2), and the letter prime is typed as an apostrophe (section 5.1). The only remaining diacritics from the Archimedes list for which we have no rule yet are the markers for short and long vowels/syllables: breve (ě), macron (q̄). Do we need to have a rule for them, or can we ignore them?

`<+>` and `<->` may be introduced for indicating different font sizes. Alternative: `<s+>`, `<s->` for superscript and subscript, `<f+>`, `<f->` for larger and smaller font sizes.

Italics

figures, example 2: Paragraph in italics, except THEON at the beginning: I have added the use of `__` in `<p it>`-paragraphs as a rule. (See also the first question in the Latin Example section.)

Problem of accidental italic letters within a paragraph in italics: “Ignore the italics of single characters within a word in normal font”: leave out? (Does not apply to mathematical texts anyway?? The standard of using italics is not so old, I guess?)

Latin Ligatures

In addition to the easy ligatures `ff`, `fi`, `fl`, `ffi`, `ffl`, `ff`, `fi`, `ffi`, `ft`, `ſt` `ct`, Hoefler Text also has glyphs for `fb`, `fh`, `fj`, `fk`, `ffb`, `ffh`, `ffk`, `fb`, `fh`, `fk`, but I guess this leads too far?

leave out the `ij` ligature?

is the `{quis}` really `quis`?

Latin example

The capital letters of the captions in italics are mostly not in italics (contrary to the capital letters in the heading in italics). Do we want to mark that. i.e.

`<cap it>_P_unctus, aut $ignum. _L_linea.</cap>`, or

`<cap>P_unctus, aut $ignum. _L_linea. </cap>` in analogy to a (hypothetical)

`<h>H<sc>EADING IN</sc> S<sc>MALL</sc> C<sc>APS</sc></h>`,

or do we simply ignore it because it has little semantic value? Will they notice it at all?

before `<?>`: is the `s`, handwritten?? If yes, by whom?


lōgītudo versus longitudinem: Is it only a possible variation that could be used to fill the lines more evenly?

Does the structure of `<cap>` and `<fig>` in the example make sense?

`<fig><fig><fig>`: According to our own rules this is not correct, but optically it makes sense. Stick with it or change it so that we obey our own rules literally? And: A sentence in the note that one could see the three figures as one figure, or even the four figures as one figure? Or would that be confusing?

Greek example

Is the Greek font big enough? Especially the diacritics may be difficult to see.

Greek whole-page example, line δ': In the word {εξ}ίσ{ου}, the {εξ}-ligature is resolved correctly in Ingram 1966, but he resolves the similar-looking  as {εξ} too, whereas the LaTeX philokalia package produces the latter with {οξ}. One of them must be wrong, unless different books use the same image for different ligatures. Check this!

I have used the apostrophe for the numbers, e.g. δ', as well as for the ellipsis sign, e.g. ἐφ'. Is that a problem? (In addition, I use it to type the prime in a mathematical text, but at least that shouldn't be a problem since the Greeks didn't work with derivations of functions.)

(Comparison with a modern edition of the Elements: ἐξ ἴσου is (twice) written as one word, and in one sentence the ἐστὶν is at a different position. Lots of commas that are missing in the modern edition.)

archimedes.pen

Extract a sensible list of strange latin characters from `archimedes.pen`: Already IN: footnote symbols ¶ (pilcrow), † (dagger); mathematical symbols ' (prime), " (double prime), ± (plusminus), ∴ (therefore). Already OUT: all spacing diacritics, □ (white rectangle), ... (horizontal ellipsis). Symbols that we have already or that don't make sense for us: (EM DASH), | (lbrk, verbar; what's the difference?), { (lcub), (NON-BREAKING SPACE), (SOFT HYPHEN), ~ (tilde), (LEFT and RIGHT PARENTHESIS for Greek texts in Beta Code), (LEFT and RIGHT SQUARE BRACKET).

It turns out that the remaining symbols are also either footnote symbols (I have added a table in section 2.4.2) or mathematical symbols (I have added a table in section 5.1).

In particular: additional IN: footnote symbols ‡ (DOUBLE DAGGER U+2021; archimedes.pen mixes this up with ↯, U+2921), § (SECTION SIGN), and an additional footnote symbol that was not in the Archimedes list; mathematical symbols ° (DEGREE SIGN), ∞ (INFINITY), ÷ (DIVISION SIGN), · (MIDDLE DOT; different rules for the mathematical symbol and the Greek punctuation mark), √ (RADICAL SIGN), and ∫ (INTEGRAL SYMBOL; I have never seen this symbol before, but it's easy to type with an escape sequence). Additional OUT: © (COPYRIGHT SIGN), ℞ (PRESCRIPTION TAKE), ⊂ (CLOSED SUBSET), ★ (BLACK STAR), ☾ (FIRST QUARTER MOON).

Appendix

Add lists of ligatures for Latin and Greek, footnote symbols and mathematical symbols? Contrary to the list of tags, there are four clearly defined places in the document where this information can be found, so the appendix would only copy information from these sections.

F To Do

F.1 To Do for Version 1.1.1

Malcolm will go over ze English. Probably the first version that goes to China.

General

Make sure that the second part of the appendix doesn't go to China.

Avoid sparsely printed pages (use `\enlargethispage`? The margins are quite small anyway. Use `\enlargethispage` with negative values?)

Make sure that the whole-page examples are on a left and right page. (Assuming that they print it in the original A4 format; if they print two pages on one, for example, it would be the other way round. But that is not our problem, I guess.) Move Appendix A (one page) behind the examples if necessary? (And back if again necessary after another revision, etc.??)

Make sure that there is no page break immediately after `EXAMPLE` etc. Try to have images and the transcribed text the same page. Try to avoid page breaks in lists and tables.

Keep track of the origins of the examples (which book, which page). Put it in comment lines in the TeX file, as well as in a separate document.

F.2 To Do for Version 1.2

Fraktur

real examples for bold face, subscript, superscript

F.3 To Do for Version 2.0

Chinese

F.4 To Do for Later Versions

Introduce shorter tag names if there are no naming collisions?

indexes (there is an example, see the wiki)

lists: find examples

Mathematics: Rules that also work for more complex formulas (i.e. as complex as the early typographers could manage; for example, no rule is needed for $\infty^{\frac{1}{2}}$, since in all likelihood it would have been set as $\infty^{\frac{1}{2}}$ anyway)

Book-specific rules

section “Book-specific Specifications”, subsections “A Scheme for Book-specific Specifications” and “An Example”

Tables: more precise rules for books with many (e.g. astronomical) tables

Headings: maybe different rules for specific books

modular structure: is it enough to say: For this specific book, ignore sections ... ?

(possible book-specific rule: space after sentence number)

F.5 Points for the Post-processing Stage

SR suggests some minimal post-processing when texts are added to the repository, namely a concordance of page image numbers and the page numbers as they appear on the page, so that the pages can be cited properly; and also checking for doubly used and missing page numbers. SR also mentioned old ECHO wish lists for the xml; where are these lists?

`<lang gr>`, etc., for example with Latin and Greek columns? If the standard language of the book is Latin, mark Greek columns only? Or is it easily insertable during post-processing? Is it needed at all, since they use unicode characters?

During post-processing, will we correct the main text according to errata lists? Or at least links? What about new typos in the errata lists? (There is one in the example!)

Is the open rule for figures, combined with some examples, sufficient? The main criticism of the archimedes style `<p>text<figure/></p>`, i.e. the figure tag within the paragraph, concerns only the post-processing stage.

When the last paragraph on a page continues on the next page, but there is a footnote at the bottom: In the post-processing stage, all notes are moved to the end of the text (with links).

The main reason for typing footnotes where they are (rather than marginal notes and LaTeX style footnotes) is that a footnote can continue on the next page. How do we encode the second part of the footnote? Ignore the problem in the specs, hoping that they will at least identify it as a (separate) paragraph, and merge the footnote parts in the post-processing stage?

Α, Β, Γ, Δ: Α and Β will inevitably be typed as Latin characters even though semantically they are Greek characters (same shape, but different Unicode codes). Are we fine with that? Problem for searching?

alt-9 on a Polytonic Greek produces the nearly-ASCII mid-dot, not the Ano Teleia. alt-8 produces the Unicode character BULLET. Be prepared that these characters have been mixed up in the text.