

Formatting Moby-Dick

Introduction	1
Successive Approximations	2
Styling from Page Images	2
Front Matter	7
Title page	7
Book title	7
Contents	9
‘Etymology’ and ‘Extracts’	11
Body	12
Chapter separator	12
Footnotes	15
Duplicate footnotes	16
Footnote size	18
Block	19
Widows and orphans	19
Hyphen at end of page	21
Text	21
Italics and small-caps	21
‘Curly’ quotes	22
Consecutive em dashes	24
Baseline grid	25
Headers and Footers	27
Conclusion	28
References	28

Introduction

This document describes aspects of the stylesheets that were developed to format the first American edition of Moby-Dick by Herman Melville. The stylesheets illustrate one way to approach developing a stylesheet for XSL-FO, and they also illustrate how to use some AH Formatter extensions.

The source for Moby-Dick (1) is TEI-encoded XML (2) from the Wright American Fiction project (3).

The stylesheets were developed for a project to demonstrate how to use the Automated Analysis feature (4) of AH Formatter V7.0 (5). As such, the initial emphasis was on getting the text block of the body pages correct. The styles for everything outside the text block – headers and footers, the front-matter, and the advertisements at the back of the first edition – were initially developed as a rough approximation of the formatting used in the first edition. Over time, the styles have been refined to more accurately mimic the printed first edition.

Successive Approximations

To develop a stylesheet for formatting with either XSL-FO or CSS is usually a process of developing successive approximations of the final result. This is true whether the look of the document is being developed on the fly, developed according to a design brief, or developed to match an existing document, as with Moby-Dick.

The first draft of a stylesheet will likely produce only a rough approximation of the final result. If you are developing on the fly, then you haven't made up your mind about the final look at that point anyway. If you are developing according to a design brief, then the first version that you format is likely to have the correct page size and the correct fonts and font sizes for major titles and paragraphs but may omit more context-specific styles such as for the table of contents, index, tables, nested lists, and so on. It is similar for developing styles to match an existing document.

That is usually followed by a sequence of making and reviewing changes to bring the styles closer to the final result. This is true, of course, when you are developing on the fly, because the final result isn't known until you say that you have the result that you want. It is also true for both developing according to a design brief and developing to match an existing document, because there are additional contexts that you know you have not handled yet and, quite likely, more contexts that neither you nor the designer had anticipated. These might include nesting lists of different types or handling figures or table immediately after a title or, for Moby-Dick, handling Queequeg's mark or stage directions and songs.

Successive changes should, of course, bring you closer to the final result. In reality, some changes will have to be redone, and some changes will throw up new problems, but the overall movement is to close in on the final result.

Styling from Page Images

The initial styles for the front-matter of Moby-Dick, in particular, were refined by setting a photograph of a page from the first edition as the background image for the corresponding page and adjusting the XSL-FO to match. The following image shows the formatted title page

Formatting Moby-Dick

with the photograph of the first edition's title page as the page background opened in the AH Formatter GUI.



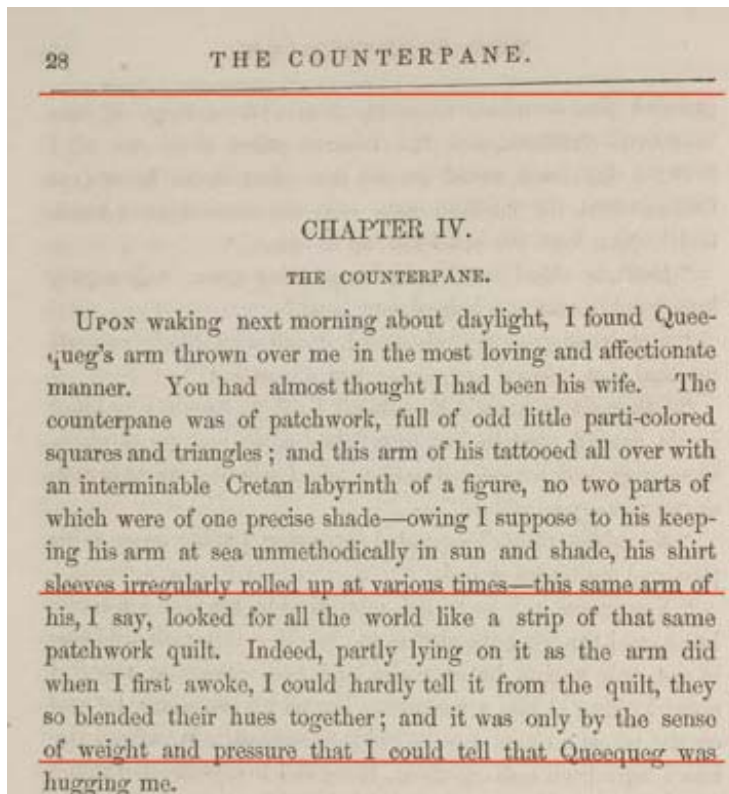
The sequence of steps to use the page image as the background is:

1. Modify a copy of the XSL-FO to add `axf:bleed` and `axf:crop-offset` properties to each `fo:simple-page-master` that will have a background image. For example:

```
<fo:simple-page-master master-name="First-PageMaster"
  page-height="7.375in"
  page-width="4.78in"
  axf:bleed="0.5in"
  axf:crop-offset="0.5in">
```

2. If necessary, rotate the page image so that the text is as horizontal as possible.

The first edition is now 170 years old, and the available page images are photographs of pages in the bound book, rather than scans of individual pages. The result is that the text in the scans is not always perfectly parallel, either because of the condition of the page or because of the curve of the paper when the page was photographed. The following image shows that variation can happen: the red lines are parallel, the text is not.



3. Specify the page image, scaled and positioned to match the formatted page, as the background image of either the fo:simple-page-master:

```
<fo:simple-page-master
  master-name="First-PageMaster"
  page-height="7.375in"
  page-width="4.875in"
  background-image="page-images/MD_Amer_0038.jpg"
  axf:background-size="5.21in"
```

```
background-position="-0.12in -0.15in"  
axf:bleed="0.5in"  
axf:crop-offset="0.5in">
```

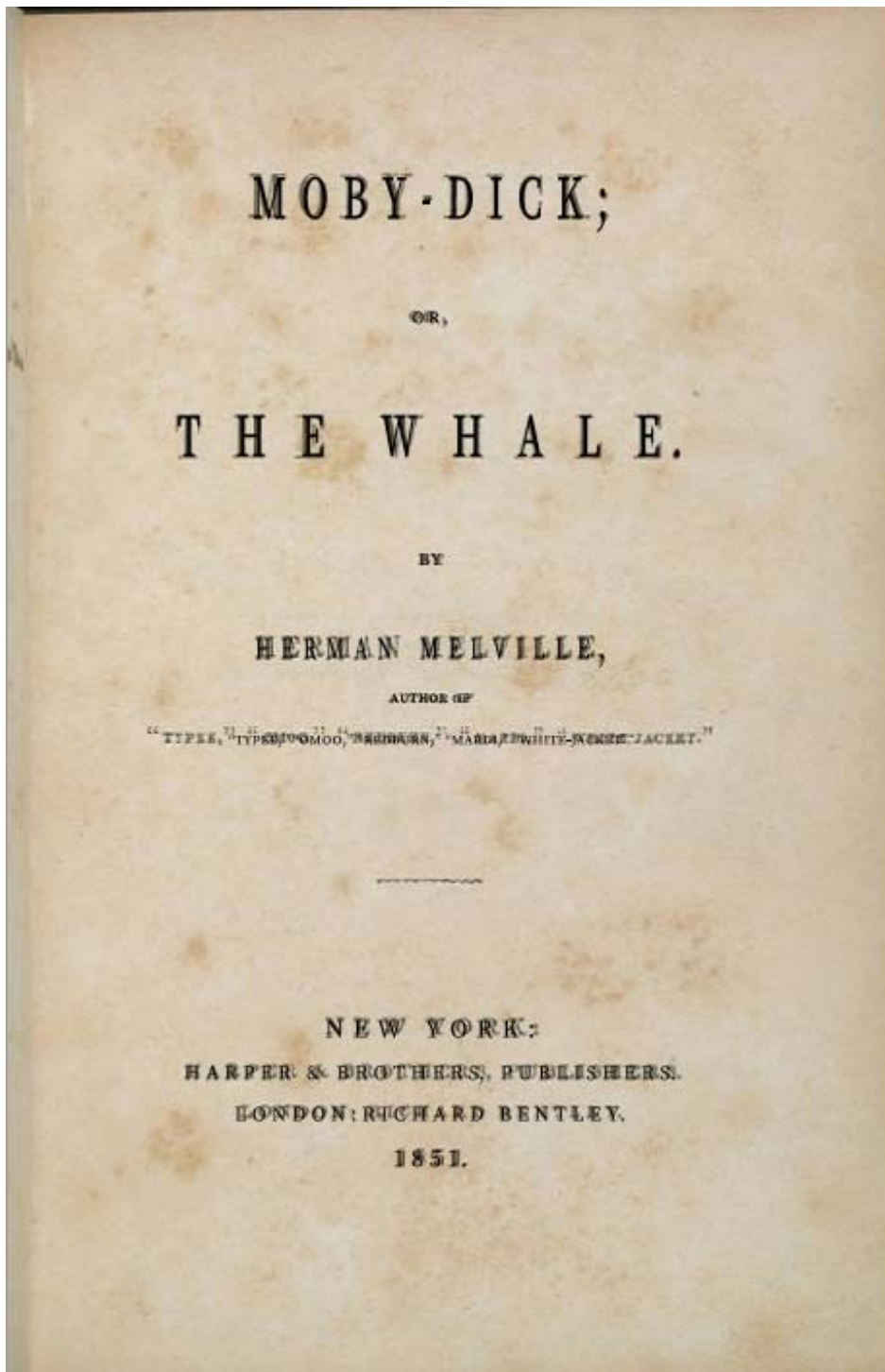
or on the fo:page-sequence that generates the page:

```
<fo:page-sequence  
  master-reference="CoverFrontMaster"  
  background-image="page-images/MD_Amer_0019.jpg"  
  axf:background-size="5.7in"  
  background-position="-0.7in -0.3in">
```

Because the page images for the first edition are photographs, there was considerable variation in the size and position of the page within each image. Getting the correct size and position was an iterative process of modifying the XSL-FO and viewing the result in the AH Formatter GUI, then repeating the process until the result is satisfactory. Enabling 'Show Borders' in the AH Formatter GUI makes it easier to judge how to adjust the background image.

4. Iteratively modify the XSL-FO then view it in the AH Formatter GUI until the formatted document satisfactorily matches the page from the first edition.
5. Modify the stylesheets for generating the XSL-FO to recreate the FOs and properties that were arrived at manually.

The result can be quite a close approximation of the original:



The different parts of the front matter of the first edition show considerable variation in fonts, font sizes, and letter- and word-spacing. That, combined with the necessarily imprecise size and position of the background images, has resulted in a range of values for the same properties applied at different places on different pages. When time permits, it should be possible to rationalize these and use fewer, more consistent values and still reproduce the first edition pages with sufficient accuracy. After all, the first edition was printed with a fixed set of

founts and with fixed increments of the space that could be added between letters. Font sizes, etc., were unlikely to have been specified in points in America in 1851, but the sizes would have been internally consistent.

Front Matter

The front matter of Moby-Dick comprises:

- Title page
- Copyright page
- Dedication
- Contents
- Fly title
- Etymology
- Extracts

Title page

As shown previously, it is possible to reproduce the title page fairly accurately.¹

Book title

The markup for the book's title does not include sufficient information to accurately reproduce the formatted title:

```
<docTitle>
  <titlePart>MOBY-DICK;</titlePart>
  <titlePart type="sub">OR, THE WHALE.</titlePart>
</docTitle>
```

plus the book's title is formatted identically on the fly title page, but its markup has even less correspondence to the formatting:

```
<div type="fly_title">
  <head>MOBY-DICK; OR, THE WHALE.</head>
</div>
```

Because the stylesheet is specific to Moby-Dick, it was simpler to ignore the markup and to use `xsl:analyze-string` and generate FOs around parts of the title text:

```
<xsl:template match="docTitle | div[@type = 'fly_title']/head"
  priority="5">
  <fo:block
    font-size="24pt"
```

¹ At the time of this writing, the formatting of the list of previous Herman Melville novels is not yet styled quite like the first edition. The quotation marks in the first edition are a larger font size than the titles. `xsl:analyze-string` will be used to add `fo:inline` elements around the quotation marks to change their font size.

```

        letter-spacing="0.37em"
        line-height="1"
        text-align="center"
        font-stretch="extra-condensed">
<xsl:analyze-string
    select="normalize-space(.)"
    regex="OR,">
<xsl:matching-substring>
    <fo:block
        font-size="8pt" font-variant="all-small-caps"
        font-stretch="normal"
        letter-spacing="0.125em" space-before="30pt">
        <xsl:value-of select="." />
    </fo:block>
</xsl:matching-substring>
<xsl:non-matching-substring>
    <fo:block axf:letter-spacing-side="start">
        <xsl:if test="contains(., 'THE WHALE.')">
            <xsl:attribute name="space-before" select="'30pt'"
/>
            <xsl:attribute name="letter-spacing"
select="'0.9em'" />
        </xsl:if>
        <xsl:analyze-string
            select="."
            regex="\.| ">
            <xsl:matching-substring>
                <fo:inline letter-spacing="0.3em">
                    <xsl:value-of select="." />
                </fo:inline>
            </xsl:matching-substring>
            <xsl:non-matching-substring>
                <xsl:value-of select="." />
            </xsl:non-matching-substring>
        </xsl:analyze-string>
    </fo:block>
</xsl:non-matching-substring>
</xsl:analyze-string>
</fo:block>
</xsl:template>

```

The document, with a few minor exceptions, is formatted entirely in Source Serif Pro. The font is both open source and a reasonable match for the font that was used for paragraphs in the first edition. However, in the first edition, the title page (and some other titles) uses both narrow and small-capital variants of the text font. Source Serif Pro does not have a narrow

Formatting Moby-Dick

version, so the narrow variants are achieved by setting the 'font-stretch' property (for example, 'font-stretch="extra-condensed"') and relying on AH Formatter to adjust each character's width. Source Serif Pro does have true small caps (6), but 'font-variant="small-caps"' as defined in XSL 1.1 uses small caps only for lower-case letters. Because the small caps in the title are represented in the XML as capital letters, it is necessary to use the 'font-variant="all-small-caps"' AH Formatter extension to format the capital letters in the source as small caps.

Many of the titles in the first edition use letter-spaced characters. Letter spacing is specified with the 'letter-spacing' property, and values such as 'letter-spacing="0.37em"' were arrived at through trial and error to match the appearance of the page image when it was used as the background. However, in the first edition, the letter spacing between an alphabetic character and a following punctuation character is sometimes less than the letter spacing between two alphabetic characters. The solution is to use xsl:analyze-string to generate an fo:inline with a different 'letter-spacing' value around just those characters. A refinement, used elsewhere in the stylesheet, is to also use the 'axf:letter-spacing-side' AH Formatter extension so that all of the space that is added is between the alphabetic characters added at their start side and so does not contribute to the space between an alphabetic character and a following punctuation character.

Contents

Chap.	Page	Chap.	Page
I.—Loomings	1	XI.—Forecastle.—Mid-	
II.—The Carpet Bag	7	night	189
III.—The Spouter-Inn	11	XII.—Moby Dick	196
IV.—The Counterpane	28	XIII.—The Whiteness of	
V.—Breakfast	32	the Whale	207
VI.—The Street	35	XIII.—Hark!	217
VII.—The Chapel	37	XIV.—The Chart	218
VIII.—The Pulpit	41	XV.—The Affidavit	224
IX.—The Sermon	44	XVI.—Surmises	234
X.—A Bosom Friend	54	XVII.—The Mat-Maker	237
XI.—Nightgown	58	XVIII.—The First Lowering	240
XII.—Biographical	61	XIX.—The Hyena	252
XIII.—Wheelbarrow	63	L.—Ahab's Boat and	
XIV.—Nantucket	69	Crew.—Fedallah.	255
XV.—Chowder	71	L.—The Spirit-Spout	258
XVI.—The Ship	75	LII.—The Pequod meets	
XVII.—The Ramadan	91	the Albatross.	262
XVIII.—His Mark	97	LIII.—The Gam.	264
XIX.—The Prophet	102	LIV.—The Town Ho's	
XX.—All Astray	106	Story	269
XXI.—Going Aboard	108	LV.—Monstrous Pictures	
XXII.—Merry Christmas	112	of Whales	292
XXIII.—The Lee Shore	117	LVI.—Less Erroneous Pic-	
XXIV.—The Advocate	118	tures of Whales	298
XXV.—Postscript	123	LVII.—Of Whales in Paint,	
XXVI.—Knights and Squires.	125	in Teeth, &c	302
XXVII.—The Spouter-Inn	128	LVIII.—Brit	305
XXVIII.—Ahab	133	LIX.—Squid	308
XXIX.—Enter Ahab; to him,		LX.—The Line	311
Stubb	137	LXI.—Stubb kills a Whale	315
XXX.—The Pipe	141	LXII.—The Dart	321
XXXI.—Queen Mab	142	LXIII.—The Crotch	322
XXXII.—Cetology	144	LXIV.—Stubb's Supper	324
XXXIII.—The Specksynder	159	LXV.—The Whale as a	
XXXIV.—The Cabin Table	162	Dish	333
XXXV.—The Mast-Head	168	LXVI.—The Shark Mas-	
XXXVI.—The Quarter-Deck	176	sacre	336
XXXVII.—Sunset	185	LXVII.—Cutting In	338
XXXVIII.—Dusk	186	LXVIII.—The Blanket	340
XXXIX.—First Night-Watch	188	LXIX.—The Funeral	343
		LXX.—The Sphynx	345

Chap.	Page	Chap.	Page
I.—Loomings	1	XI.—Forecastle.—Mid-	
II.—The Carpet Bag	7	night	188
III.—The Spouter-Inn	11	XII.—Moby Dick	196
IV.—The Counterpane	28	XIII.—The Whiteness of	
V.—Breakfast	32	the Whale	207
VI.—The Street	35	XIII.—Hark!	217
VII.—The Chapel	37	XIV.—The Chart	218
VIII.—The Pulpit	41	XV.—The Affidavit	224
IX.—The Sermon	44	XVI.—Surmises	234
X.—A Bosom Friend	54	XVII.—The Mat-Maker	237
XI.—Nightgown	58	XVIII.—The First Lowering	240
XII.—Biographical	60	XIX.—The Hyena	252
XIII.—Wheelbarrow	63	L.—Ahab's Boat and	
XIV.—Nantucket	68	Crew.—Fedallah.	255
XV.—Chowder	71	L.—The Spirit-Spout	258
XVI.—The Ship	74	LII.—The Pequod meets	
XVII.—The Ramadan	90	the Albatross	262
XVIII.—His Mark	97	LIII.—The Gam	264
XIX.—The Prophet	101	LIV.—The Town Ho's	
XX.—All Astray	105	Story	269
XXI.—Going Aboard	108	LV.—Monstrous Pictures	
XXII.—Merry Christmas	112	of Whales	292
XXIII.—The Lee Shore	117	LVI.—Less Erroneous Pic-	
XXIV.—The Advocate	118	tures of Whales	298
XXV.—Postscript	123	LVII.—Of Whales in Paint,	
XXVI.—Knights and Squires	124	in Teeth, &c	302
XXVII.—Ahab	133	LVIII.—Brit	305
XXVIII.—Enter Ahab; to him,		LIX.—Squid	308
Stubb	137	LX.—The Line	311
XXX.—The Pipe	140	LXI.—Stubb kills a Whale	315
XXXI.—Queen Mab	141	LXII.—The Dart	321
XXXII.—Cetology	144	LXIII.—The Crotch	322
XXXIII.—The Specksynder	159	LXIV.—Stubb's Supper	324
XXXIV.—The Cabin Table	162	LXV.—The Whale as a	
XXXV.—The Mast-Head	168	Dish	333
XXXVI.—The Quarter-Deck	176	LXVI.—The Shark Mas-	
XXXVII.—Sunset	184	sacre	336
XXXVIII.—Dusk	186	LXVII.—Cutting In	338
XXXIX.—First Night-Watch	187	LXVIII.—The Blanket	340
		LXIX.—The Funeral	343
		LXX.—The Sphynx	345

The Table of Contents is formatted in two columns. It is marked up as a list but is rendered as a four-column table to be able to recreate the formatting of the first edition.

The TEI for the Table of Contents begins:

```
<div type="contents">
  <pb n="v (Table of Contents) " xml:id="VAC7237-00000003"/>
  <head>CONTENTS.</head>
  <list>
    <item>I.–Loomings. <ref target="VAC7237-00000013"
rend="right">1</ref>
    </item>
    <item>II.–The Carpet Bag. <ref target="VAC7237-00000017"
rend="right">7</ref>
    </item>
```

The content of the <ref> elements is each chapter's page number in the first edition. The 'target' attribute, however, refers to <pb> milestone elements that mark the start of each two-page spread. At least one of the cross-references was found to point to the spread after the first page of its chapter and had to be corrected. More may yet be found.

The cross-references could not be used anyway because the XSL-FO version does not attempt to recreate the page breaks of the first edition. The cross-references from Table of Contents entries to chapters in the generated PDF are determined from the position of the list item for each chapter in the Table of Contents list:

```
<!-- Every chapter has a generated ID, and 'EPILOGUE.' is the
only
  ToC entry without a page number. -->
<xsl:variable
  name="target"
  select="if (exists(ref))
    then concat('chapter-', position())
    else 'epilogue'"
  as="xs:string" />
```

The markup for each chapter begins '<div type="chapter">', but generating the ID for each chapter could not use 'position()' because some of the <pb> milestones appear between chapters:

```
<fo:block
  id="{@type}-{count(preceding::div[@type = current()/@type])
+ 1}">
```

The Table of Contents is formatted as a four-column table to keep the different parts of the Table of Contents entries aligned:

- Chapter number (in small-caps roman numerals)
- Em-dash
- Chapter title and leader dots

The alignment and spacing of the leader dots is simple with XSL-FO:

```
<fo:leader leader-pattern="dots"
           leader-pattern-width="1em"
           leader-alignment="end" />
```

- Page number

‘Etymology’ and ‘Extracts’

The ‘Etymology’ and ‘Extracts’ segments each consist of an introductory narrative page followed by quotes and, in ‘Etymology’, a table. The fonts used for the titles in the first edition are not consistent, so they each needed a separate template.

ETYMOLOGY.

ETYMOLOGY

EXTRACTS.

EXTRACTS.

In both ‘Etymology’ and ‘Extracts’, each quote has an attribution, and each attribution is marked up as following the quoted material:

```
<cit>
  <q>
    <p>"Very like a whale."</p>
  </q>
  <bibl>
    <title>Hamlet</title>.</bibl>
</cit>
```

However, if there is enough space on the last line, the attribution is formatted in the same <fo:block> as the quotation:

“Very like a whale.”

Hamlet.

If there is not enough space, because either the last line or the attribution is too long, then the attribution is formatted on the next line.

Placing the attribution in the same block is handled by the common XSLT pattern of not formatting the <bibl> as part of the default processing and instead explicitly selecting its content when processing the <q>:

```
<xsl:template match="q/p">
  <fo:block>
    <xsl:apply-templates />
    <xsl:if test="position() = last() and
      exists(..//following-
sibling::*[1][self::bibl])">
      <fo:leader leader-pattern="space"/>
      <fo:leader leader-pattern="space"
        leader-length.optimum="100%"/>
      <fo:inline-container padding-left="2em"
        padding-right="0.125in"
        max-width="80%" text-indent="0">
        <fo:block text-align="right">
          <xsl:apply-templates
            select="..//following-sibling::*[1]/node()" />
        </fo:block>
      </fo:inline-container>
    </xsl:if>
  </fo:block>
</xsl:template>

<xsl:template
  match="bibl[exists(preceding-sibling::*[1][self::q[p]])]"
  priority="5" />
```

Placing the attribution either on the last line of the quotation or on the next line is handled by the common XSLT pattern of using two <fo:leader>.

Body

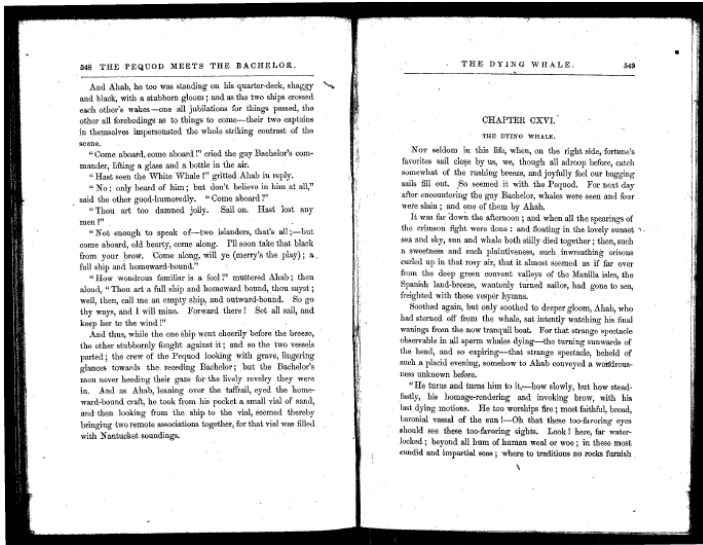
The majority of Moby-Dick is 135 chapters of largely text. Melville scholars like to find patterns in the structure of the chapters (7), but when formatting Moby-Dick, the most useful distinctions are between paragraph-like blocks of text and other content.

The non-paragraph content includes:

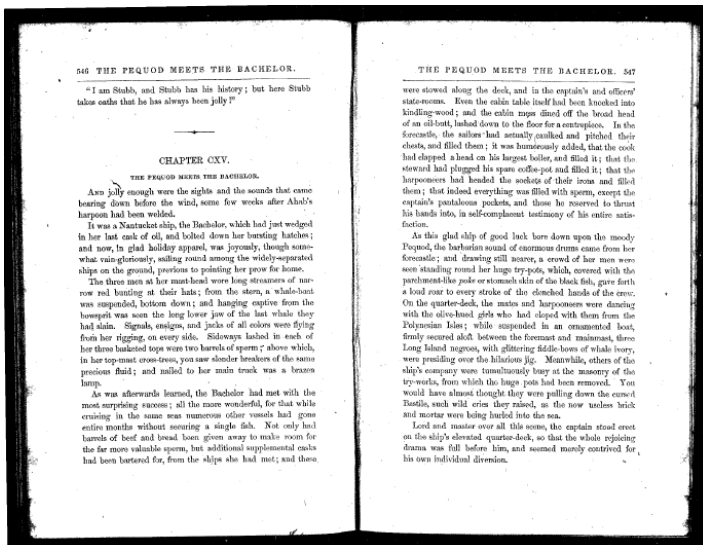
- A single graphic (for Queequeg's mark)
- Inscriptions from tombstones
- Songs and poems
- Speeches and stage directions as if for a play

Chapter separator

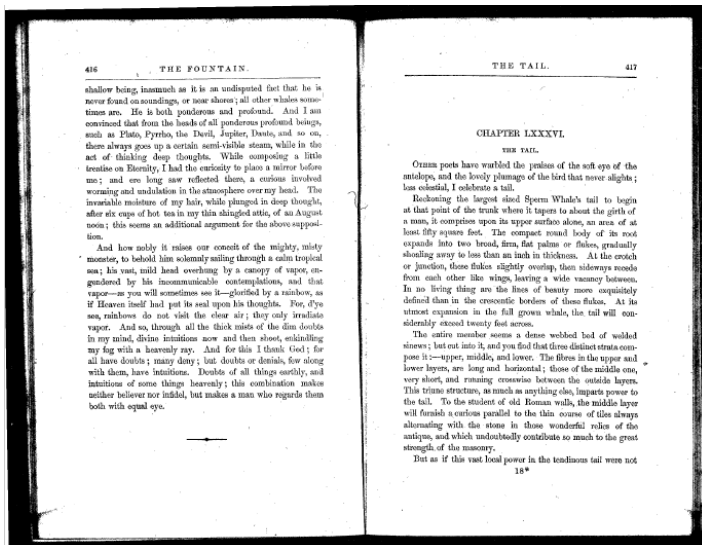
When a chapter in the first edition ends near the bottom of a page, the next chapter begins on the following page with space before the chapter title:



When a chapter does not end near the bottom of a page, there is an additional separator printed before the chapter title. To complicate matters, the space between the separator and the chapter title is less than the space before a chapter that starts on a new page:



When a chapter ends with some space left at the bottom of a page but not enough space for the separator and the chapter title, the separator is printed at the end of the chapter and the next chapter starts on the next page:



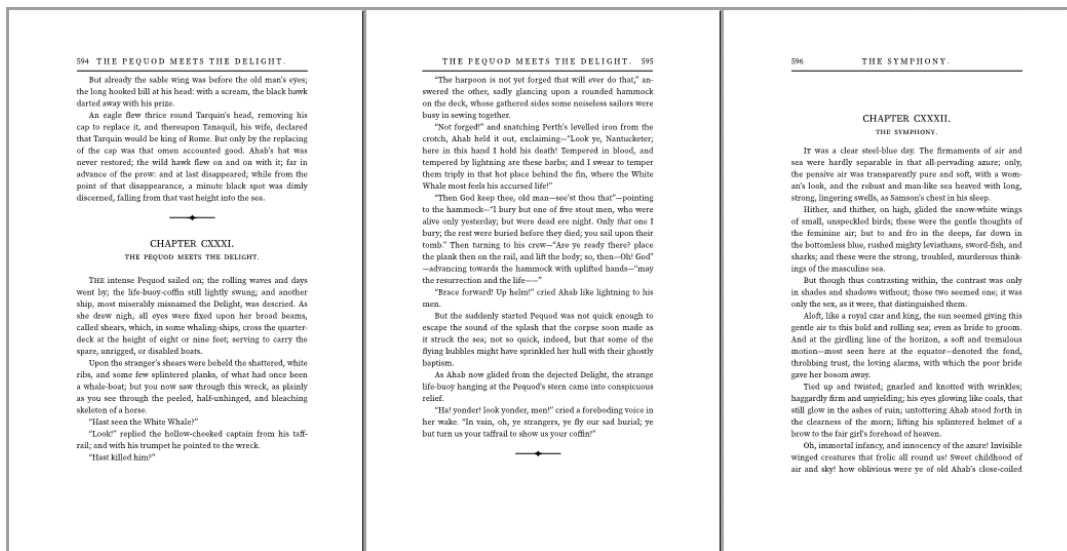
When the first edition was composed manually, it would have been straightforward to add the separator when and where it was needed. It is not quite as straightforward with automated, 'lights-out' formatting using XSL-FO. Because the page breaks are not known before the document is formatted, it is not possible to just insert as many separators as needed, and the XSL 1.1 Recommendation does not support conditional processing based on an area's position on the page.

Two things make this possible with AH Formatter: firstly, the 'axf:suppress-if-first-on-page' extension property makes AH Formatter suppress the separator for a chapter title at the top of a page; and, secondly, the standard 'space-after.precedence="force"' on the <fo:block> for the separator ensures the correct distance between the separator and the chapter title when the separator is present while allowing the different 'space-before' value on the chapter title to apply when the separator has been suppressed or is on the previous page.

```
<xsl:template
  match="div[@type = 'chapter'] [exists(head[@type = 'sub'] |
fw[@type = 'head'])]">
  <xsl:if test="exists(preceding-sibling::div[@type =
current()/@type])">
    <fo:block axf:suppress-if-first-on-page="true" text-
align="center"
      padding-top="0.125in"
      space-after="0.2in" space-after.precedence="force"
      axf:baseline-grid="none"
      axf:baseline-block-snap="none">
      <fo:external-graphic src="images/separator.svg" />
    </fo:block>
  </xsl:if>
  <fo:block
    id="{@type}-{count(preceding::div[@type =
```

Formatting Moby-Dick

```
current()/@type]] + 1}]>
  <fo:marker marker-class-name="Chapter-Title">
    <xsl:apply-templates
      select="(fw[@type = 'head'], head[@type =
'sub'] )][1]/node()"
      mode="marker" />
  </fo:marker>
  <fo:block-container
    axf:baseline-grid="none"
    axf:baseline-block-snap="none"
    keep-together.within-page="always"
    keep-with-next.within-page="always"
    space-before="{if (exists(preceding::div[1][@type =
'chapter'] ) )
      then '0.5in'
      else '0.72in'}"
    space-before.conditionality="retain">
    <xsl:apply-templates select="head" />
  </fo:block-container>
  <xsl:apply-templates select="* except head" />
</fo:block>
</xsl:template>
```



Footnotes

Footnotes are marked up as a <ref> containing the footnote marker that refers to the separate <note> containing the footnote content:

```
<p>
  <emph>Whaling not respectable?</emph> Whaling is imperial! By
old English
  statutory law, the whale is declared "a royal fish."<ref
```

```
rend="super"
  target="#note_001" xml:id="return_001">*</ref>
  <note place="foot" xml:id="note_001">
    <p><ref target="#return_001">*</ref>See subsequent chapters
for something more on this head.</p>
  </note>
</p>
```

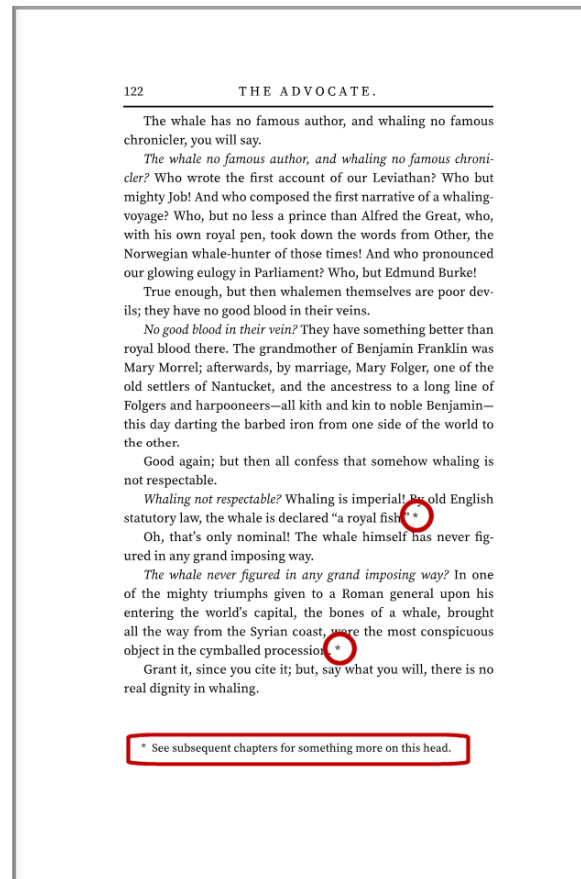
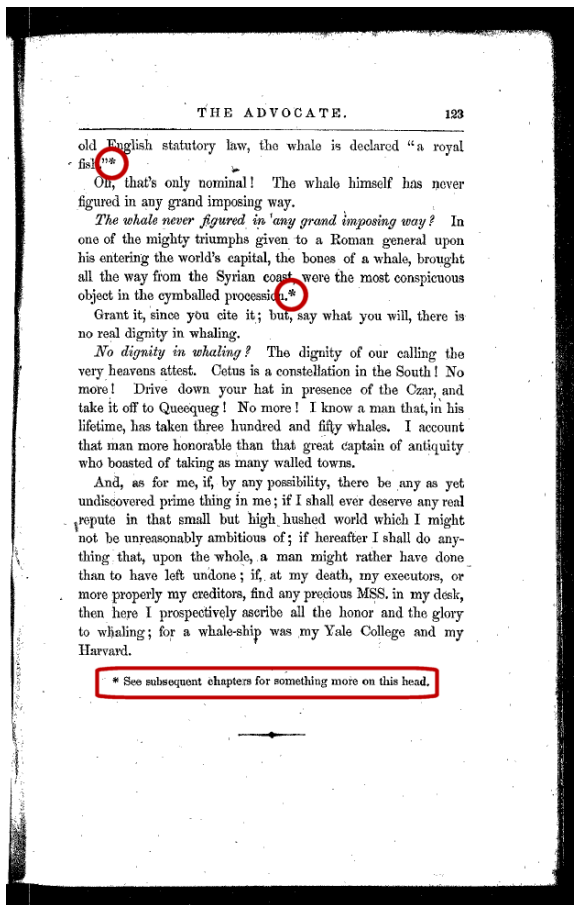
The XSL-FO `<fo:footnote>` contains both an `<fo:inline>` for the footnote marker and an `<fo:footnote-body>` for the footnote content, so the XSLT stylesheet does not process the `<note>` where it occurs in the document but instead formats the content of the `<note>` by using 'key()' to find the note that is referred to by each `<ref>`:

```
<xsl:template match="note[@place = 'foot']" />

<xsl:template match="ref[exists(key('footnote',
                                   substring-after(@target,
                                   '#')))]"
               priority="5">
  <fo:footnote
    id="{@xml:id}"
    axf:suppress-duplicate-footnote="true">
    <fo:inline>
      <fo:basic-link
        internal-destination="{substring-after(@target,
        '#')}">
        <xsl:value-of select="." />
      </fo:basic-link>
    </fo:inline>
    <fo:footnote-body
      id="{substring-after(@target, '#')}"
      font-size="7pt"
      line-height="10pt">
      <xsl:apply-templates
        select="key('footnote',
                   substring-after(@target, '#'))/node()" />
    </fo:footnote-body>
  </fo:footnote>
</xsl:template>
```

Duplicate footnotes

One page of the first edition has two references to the same footnote:



The TEI XML repeats the footnote text:

```

<p>
  <emph>Whaling not respectable?</emph> Whaling is imperial! By
  old English
  statutory law, the whale is declared "a royal fish."<ref
    rend="super" target="#note_001" xml:id="return_001">*</ref>
  <note place="foot" xml:id="note_001">
    <p><ref target="#return_001">*</ref>See subsequent chapters
    for something more on this head.</p>
  </note>
</p>
...
<p>
  <emph>The whale never figured in any grand imposing
  way?</emph> ... cymballed procession. <ref rend="super"
    target="#note_002" xml:id="return_002">*</ref>
</p>
<note place="foot" xml:id="note_002">
  <p><ref target="#return_002">*</ref>See subsequent chapters

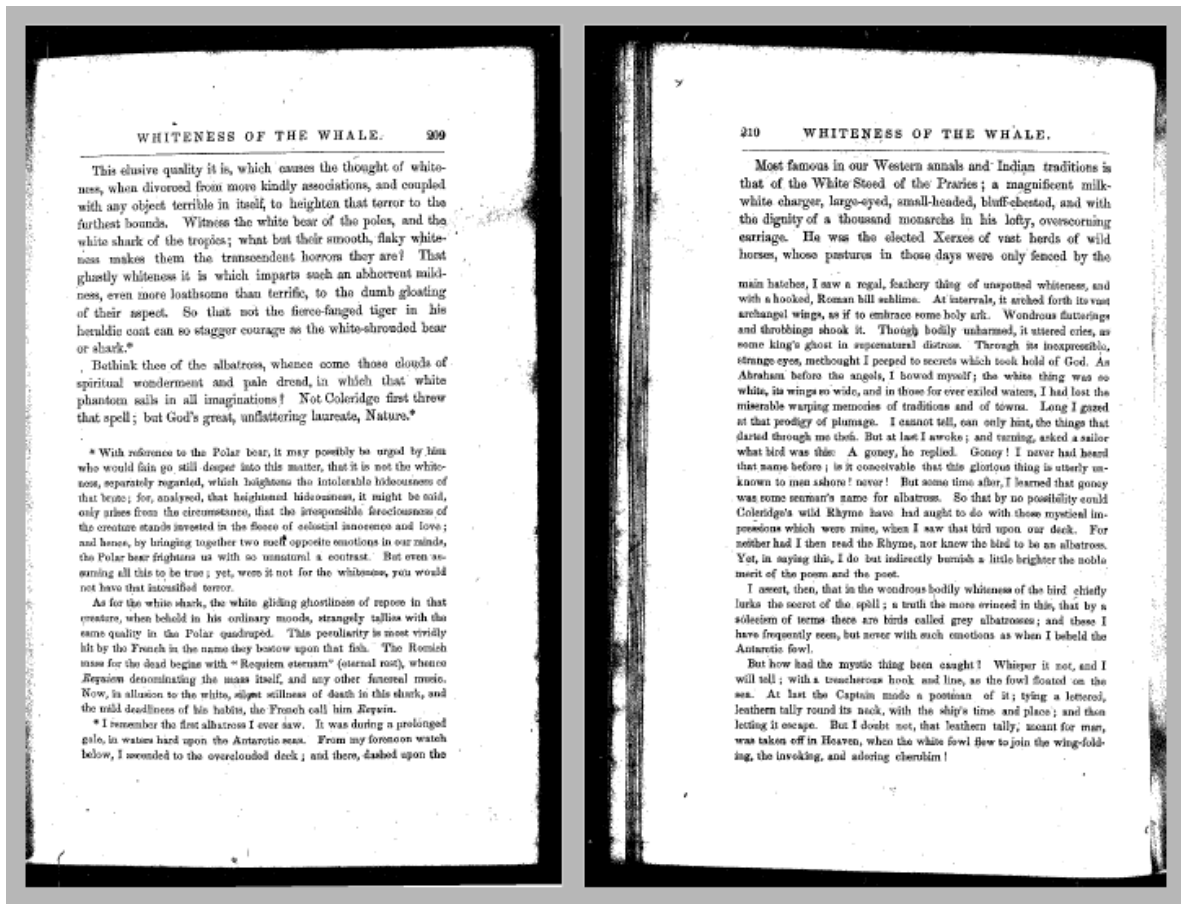
```

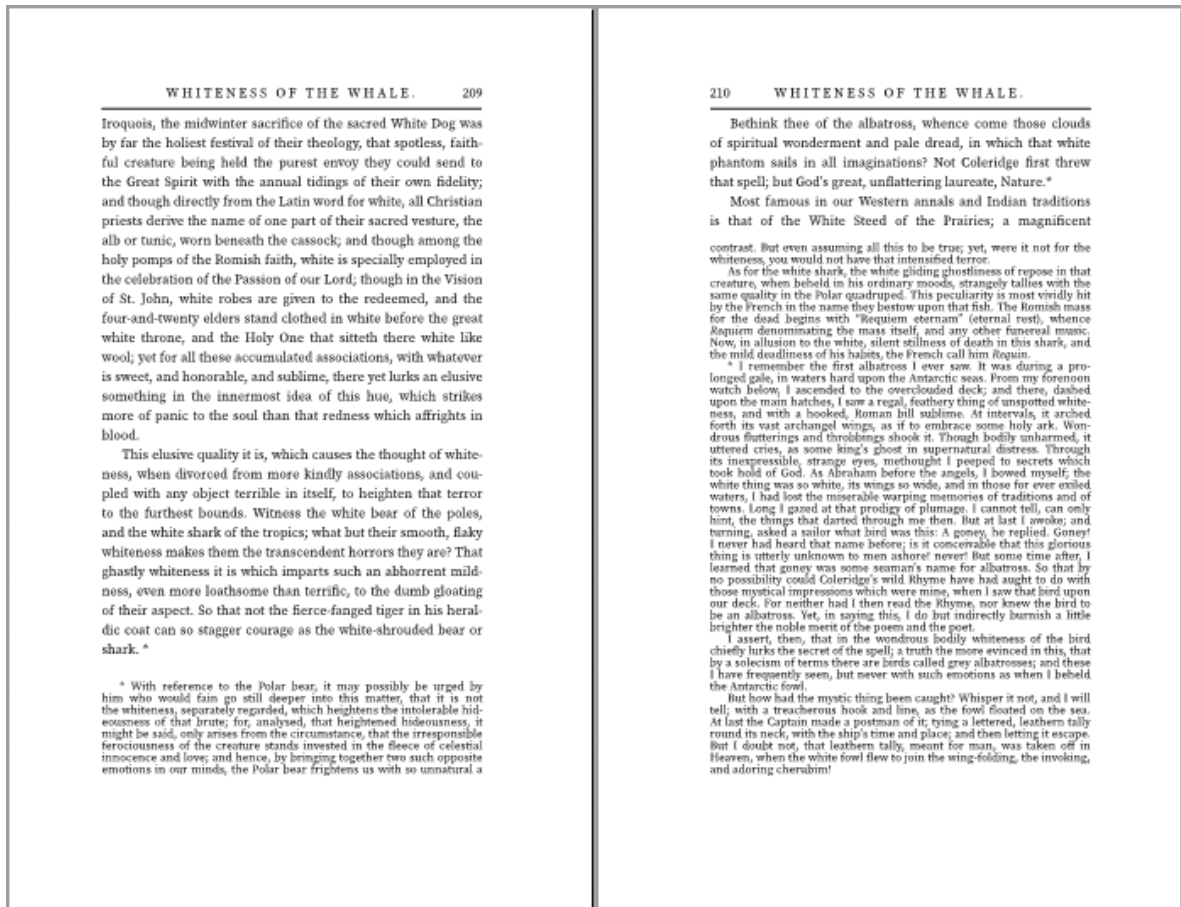
for something more on this head.</p>
</note>

A strict XSL 1.1 implementation would render both footnotes, but the 'axf:suppress-duplicate-footnote' extension property causes AH Formatter to generate only one copy of the footnote when both footnotes occur on the same page.

Footnote size

Moby-Dick also includes some whale-size footnotes:





Some things that could have been done were not needed:

- In the first edition, these two footnotes start on the same page, and it is the second footnote that continues onto a second page. Even so, both footnotes have the same ‘*’ footnote marker in the first edition.

Because the markers in the first edition are all the same, it is not necessary to use the <axf:footnote-number> and <axf:footnote-number-citation> extension elements to generate and use a sequence of footnote markers.

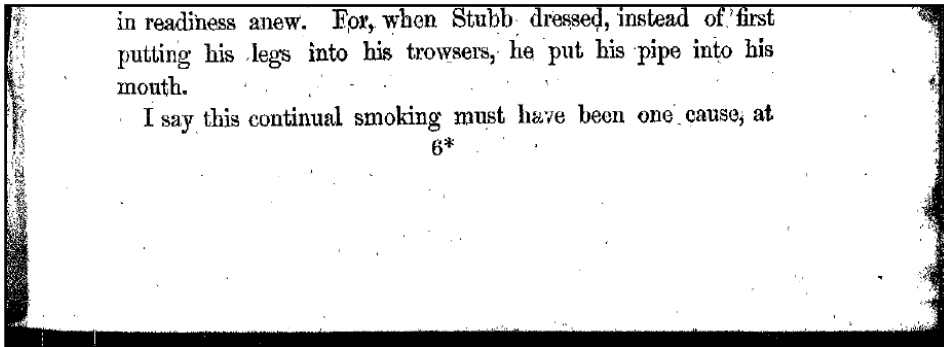
- It is possible to limit the height of the footnotes using ‘axf:footnote-max-height’, but the height of the formatted footnotes is comparable to the height in the first edition, so this also was not necessary.

Block

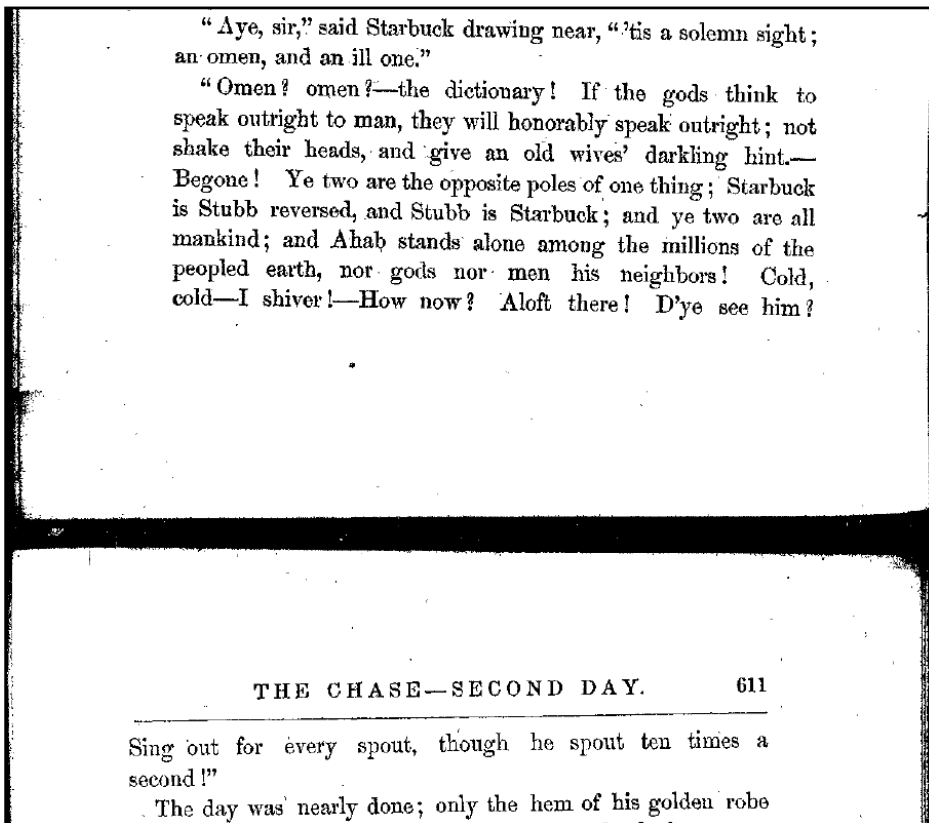
Widows and orphans

An orphan is too few lines before a page break, and a widow is too few lines after a page break.

The First Edition has multiple single-line orphans.



However, the only single lines at the top of a page are single-line dialogue. It is impossible to say how many two-line widows were deliberately forced. For example, page 610 ends with widely-spaced text, and page 611 begins with the last two lines of the paragraph:



Some of the wide spacing is due to the white-space before '?' and '!' in the First Edition, but compare the First Edition with the fewer lines when the paragraph is formatted:

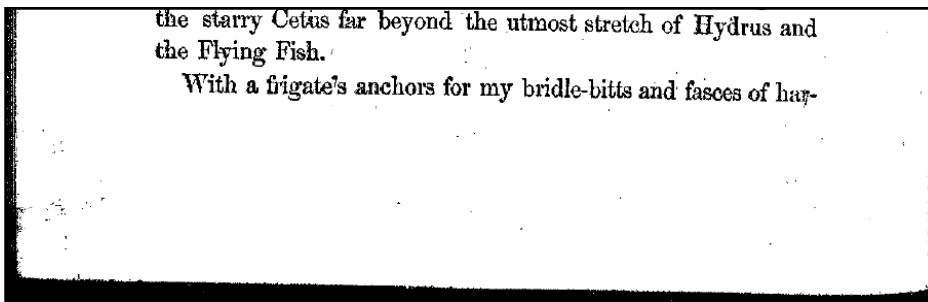
“Omen? omen?—the dictionary! If the gods think to speak outright to man, they will honorably speak outright; not shake their heads, and give an old wives’ darkling hint.—Begone! Ye two are the opposite poles of one thing; Starbuck is Stubb reversed, and Stubb is Starbuck; and ye two are all mankind; and Ahab stands alone among the millions of the peopled earth, nor gods nor men his neighbors! Cold, cold—I shiver!—How now? Aloft there! D’ye see him? Sing out for every spout, though he spout ten times a second!”

Similarly, the paragraph on pages 371–373 in the first edition is 26 lines, but is 25 lines when formatted on one page by AH Formatter. The first four formatted lines are identical to the First Edition, but then they diverge.

The formatted version uses the XSL-FO 1.1 defaults of orphans="2" and widows="2".

Hyphen at end of page

The First Edition has multiple pages that end with a hyphen:



The formatted version specifies hyphenation-keep="page" on fo:root so that words are not hyphenated across a page break. The ‘hyphenation-keep-mode’ setting in the Option Setting File is not overridden, so AH Formatter pushes only the otherwise-hyphenated word to the next page, not the entire last line.

Text

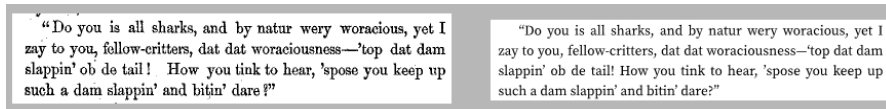
Italics and small-caps

The markup for text in italics and in small-caps needed to be corrected for proper formatting. In the original TEI XML, italic text was marked by an empty <hi rend="i"/> element at the start of the italic text but there was no indication where the italic text ended. It might be argued that to not enclose the italic text makes textual analysis easier, but foreign words (or words thought to be foreign) were marked up with a start and an end tag, for example: <foreign xml:lang="LAT">Folio</foreign>.

Text in small-caps in the first edition was included in the TEI XML as capital letters without any extra markup. It was necessary to find the text that should be small-caps, add markup, and change the text to mixed-case. For example, ‘THE’ becomes ‘<hi rend="small-caps">The</hi>’.

'Curly' quotes

Moby-Dick makes extensive use of both single- and double-quotes. This includes apostrophes replacing letters in broken English for speech from non-native speakers of English. In the first edition, the left and right quotes are visibly different:



In the TEI source XML, however, all quotes are straight:

```
<p>"Do you is all sharks, and by natur wery voracious, yet I zay to you, fellow-critters, dat dat voraciousness—'top dat dam slappin' ob de tail! How you tink to hear, 'spose you keep up such a dam slappin' and bitin' dare?"</p>
```

Converting the straight quotes to 'curly' quotes initially seemed straightforward, but it was made complicated by quotes before emphasized text and the difference between left single quotes at the start of quoted text and right single quotes at the start of a word to indicate a dropped letter.

```
<!-- Convert single and double quotes to 'curly' quotes. -->
<xsl:template match="text()" name="ahf:text">
  <xsl:param name="text" select="." as="text()" />

  <xsl:value-of select="ahf:text($text)" />
</xsl:template>

<xsl:function name="ahf:text" as="xs:string">
  <xsl:param name="text" as="text()" />

  <!-- The replacement that depends on the current node must be
  first. -->
  <xsl:variable
    name="text"
    select="if (matches($text, '&quot;') and
              empty($text/following-sibling::node()))
            then replace($text, '&quot;', '&rdquo;')
            else $text"
    as="xs:string" />
  <!-- Moby-Dick uses broken English for speech from non-native
  speakers of English. The speech can include words with
  the
  dropped initial vowel indicated by a right single-quote.
  Handle those before replacing any &apos; with left
```

```

        single-quotes. -->
<xsl:variable
  name="text"
  select="replace($text,
'' '(s?t?(&quot;|\s|[.,;:]|(balm|dent|em|gain|ll|mong|parm
|quid|specially|spose|stead|teak|till|[Tt]is|[Tt]was)(,|\s)|$
))', '&rsquo;$1')"
  as="xs:string" />
<xsl:variable
  name="text"
  select="replace($text, ' (^|\s|&quot;|-)'' (^&quot;]|$)',
'$1&lsquo;$2')"
  as="xs:string" />
<xsl:variable
  name="text"
  select="replace($text, ' (^|-|\s)&quot;', '$1&ldquo;')"
  as="xs:string" />
<xsl:variable
  name="text"
  select="replace($text, '&quot;(\s|[-.,;:]|$)',
'&rdquo;$1')"
  as="xs:string" />
<xsl:variable
  name="text"
  select="replace($text, ' ([^\s])'' ([^\s])', '$1&rsquo;$2')"
  as="xs:string" />

<!-- Variations on '* * *' in 'Extracts'. -->
<xsl:variable
  name="text"
  select="replace($text, ' \*', '&#xA0;&#xA0;*)'"
  as="xs:string" />
<xsl:variable
  name="text"
  select="replace($text, '\* ', '*&#xA0;&#xA0;')'"
  as="xs:string" />

<xsl:sequence select="$text" />
</xsl:function>

```

The 'ahf:text()' XSLT function is also used in other contexts; for example:

```

<xsl:template match="div[@type = 'fly_title']/bibl">
  <fo:block text-align="center" hyphenate="false" font-
size="5pt"
      line-height="10.5pt"

```

```
        space-before="2.33in"
        space-before.conditionality="retain">
<!-- Provide structure that is not in the source XML. -->
<xsl:analyze-string select="ahf:text(edition/text())"
                    regex="HERMAN MELVILLE,">
...

```

Consecutive em dashes

The First Edition uses two or three consecutive em dashes as a typographic effect in multiple places, for example:

——— “There Leviathan,
Hugest of living creatures, in the deep

Most typography books that cover the em dash recommend a thin space before and after the dash. For example, *Correct Composition* (8) states:

As the dash entirely fills the body sideways, it should have before and after it a thin space to prevent the interference with adjoining characters.

Many digital fonts preserve the letterpress practice that the em dash completely fills its width. However, *Source Serif Pro* includes built-in white-space before and after the stroke. This is generally useful, but it looks bad when there are consecutive em dashes:

— — “There Leviathan,
Hugest of living creatures, in the deep

It looked for a time that it would be necessary to wrap consecutive em dashes with `<fo:wrapper font-family="serif">` to select a font with em dash that would join up. However, a chance (re)discovery of the Unicode characters for two and three consecutive em dashes provided the way to show the correct dashes without changing fonts. More steps were added to the text handling:

```
<xsl:variable
  name="text"
  select="replace($text, '&mdash;&mdash;&mdash;', '&#x2E3B;')"
  as="xs:string" />
<xsl:variable
  name="text"
  select="replace($text, '&mdash;&mdash;', '&#x2E3A;')"
  as="xs:string" />

```

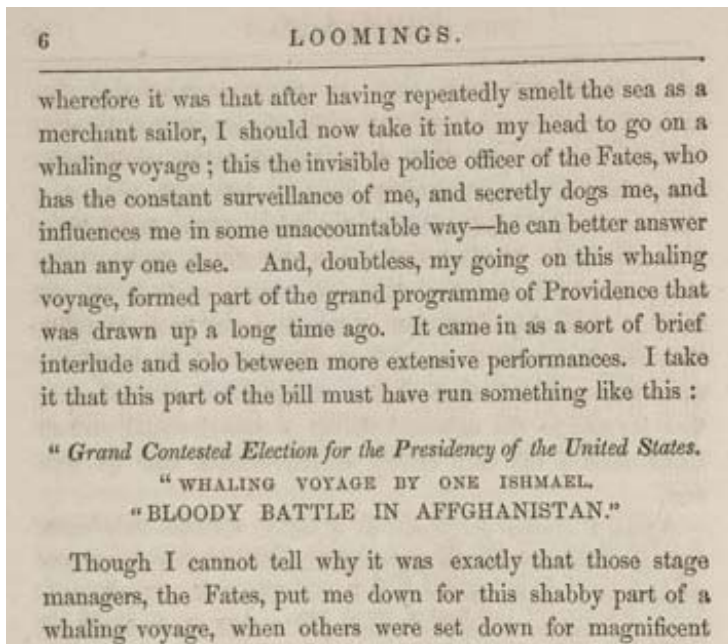
It is now possible to use the correct dashes from the same font:

— “There Leviathan,
Hugest of living creatures, in the deep

Baseline grid

‘Show-through’ occurs when text on the back of a page is visible through the paper. The shadow of the text on the back reduces the legibility of the text on the front. One way to reduce show-through (aside from using thicker paper or only reading the document electronically) is to align the lines of text on the front and back of the page.

This image from the first edition shows some show-through, but it also shows both that the lines mostly line up and that lines resume their alignment after the three irregular lines:



Keeping lines aligned front-and-back is straightforward when all of the text is the same font size and has the same line height. It becomes harder when the text includes titles, etc., that have different font sizes, line heights, and space before and after. It is often possible to style a title such that the space before the title, the line height of the title, and the space after the title add up to a multiple of the base line height. However, this will fail if some titles extend over two lines and the line height of the title is not a multiple of the base line height.

The AH Formatter baseline grid extension can both align lines to a common baseline and allow lines in specific blocks to either align to their own grid or align to no grid at all. The red lines in the following figure highlight that lines in ordinary paragraphs are aligned to the baseline grid even after the three irregular lines and after a chapter number and title:

Formatting Moby-Dick

<p style="text-align: center;">6 LOOMINGS.</p> <p>voyage; this the invisible police officer of the Fates, who has the constant surveillance of me, and secretly dogs me, and influences me in some unaccountable way—he can better answer than any one else. And, doubtless, my going on this whaling voyage, formed part of the grand programme of Providence that was drawn up a long time ago. It came in as a sort of brief interlude and solo between more extensive performances. I take it that this part of the bill must have run something like this:</p> <p style="text-align: center;"><i>"Grand Contested Election for the Presidency of the United States. "WHALING VOYAGE BY ONE ISHMAEL. "BLOODY BATTLE IN AFGHANISTAN."</i></p> <p>Though I cannot tell why it was exactly that those stage managers, the Fates, put me down for this shabby part of a whaling voyage, when others were set down for magnificent parts in high tragedies, and short and easy parts in genteel comedies, and jolly parts in farces—though I cannot tell why this was exactly, yet, now that I recall all the circumstances, I think I can see a little into the springs and motives which being cunningly presented to me under various disguises, induced me to set about performing the part I did, besides cajoling me into the delusion that it was a choice resulting from my own unbiased freewill and discriminating judgment.</p> <p>Chief among these motives was the overwhelming idea of the great whale himself. Such a portentous and mysterious monster roused all my curiosity. Then the wild and distant seas where he rolled his island bulk; the undeliverable, nameless perils of the whale; these, with all the attending marvels of a thousand Patagonian sights and sounds, helped to sway me to my wish. With other men, perhaps, such things would not have been inducements; but as for me, I am tormented with an everlasting itch for things remote. I love to sail forbidden seas, and land on barbarous coasts. Not ignoring what is good, I am quick to perceive a horror, and could still be social with it</p>	<p style="text-align: center;">THE CARPET-BAG. 7</p> <p>—would they let me—since it is but well to be on friendly terms with all the inmates of the place one lodges in.</p> <p>By reason of these things, then, the whaling voyage was welcome; the great flood-gates of the wonder-world swung open, and in the wild conceits that swayed me to my purpose, two and two there floated into my inmost soul, endless processions of the whale, and, mid most of them all, one grand hooded phantom, like a snow hill in the air.</p> <p style="text-align: center;">◆</p> <p style="text-align: center;">CHAPTER II. THE CARPET-BAG.</p> <p>I STUFFED a shirt or two into my old carpet-bag, tucked it under my arm, and started for Cape Horn and the Pacific. Quitting the good city of old Manhatto, I duly arrived in New Bedford. It was on a Saturday night in December. Much was I disappointed upon learning that the little packet for Nantucket had already sailed, and that no way of reaching that place would offer, till the following Monday.</p> <p>As most young candidates for the pains and penalties of whaling stop at this same New Bedford, thence to embark on their voyage, it may as well be related that I, for one, had no idea of so doing. For my mind was made up to sail in no other than a Nantucket craft, because there was a fine, boisterous something about everything connected with that famous old island, which amazingly pleased me. Besides though New Bedford has of late been gradually monopolizing the business of whaling, and though in this matter poor old Nantucket is now much behind her, yet Nantucket was her great original—the Tyre of this Carthage;—the place where the first dead American whale was stranded. Where else but from Nantucket did those aboriginal whalers, the Red-Men, first sailly out in canoes to</p>
---	--

The first step is to specify the baseline grid using 'axf:baseline-grid':

```
<xsl:template match="body">
  <fo:page-sequence
    master-reference="PageMaster"
    writing-mode="from-page-master-region()"
    initial-page-number="1"
    axf:baseline-grid="root">
    <xsl:call-template name="PageMaster-static-content" />
    <fo:flow flow-name="xsl-region-body" hyphenate="true"
      text-align="justify">
      <xsl:apply-templates />
    </fo:flow>
  </fo:page-sequence>
</xsl:template>
```

The second step is for blocks that do not use the baseline grid to establish their own grid, also using 'axf:baseline-grid':

```
<xsl:template match="body//q">
  <fo:block text-align="center"
    text-indent="0"
    space-before="0.25lh"
    font-size="7pt">
```

```

        line-height="9pt"
        axf:baseline-block-snap="before margin-box"
        axf:baseline-grid="new">
    <xsl:apply-templates />
</fo:block>
</xsl:template>

```

‘axf:baseline-block-snap’ specifies how a block aligns with the baseline grid, if any, of its parent block.

Headers and Footers

The headers and footers in the first edition, when present, are quite simple: just the page number and the chapter title. However, an abbreviated title is used for some chapters, even for chapters that do not have a long title. The TEI XML did not include the running header text, so any abbreviated titles were added as <fw> (“forme work”) elements (8). For example:

```

<div type="chapter">
  <head>CHAPTER XXIX.</head>
  <head type="sub">ENTER AHAB; TO HIM, STUBB.</head>
  <fw type="head" place="top-centre">ENTER AHAB.</fw>

```

It is simple to choose the <fw> element, if present, in preference to the title text as the content of the <fo:marker> for the running header:

```

<fo:marker marker-class-name="Chapter-Title">
  <xsl:apply-templates
    select="(fw[@type = 'head'], head)[1]/node()"
    mode="marker" />
</fo:marker>

```

The abbreviated title is ordinarily centered in the header:

THE VIRGIN.	401
“don't be in such a devil of a hurry to sink! By thunder,	

However, even the abbreviated title can be quite long. At least one title is long enough that it cannot be centered in the header without crowding the page number:

THE HONOR AND GLORY OF WHALING.	405
up by a whale; still, whether that strictly makes a whaleman	

The solution is to let the header overflow when it is too wide and to specify ‘axf:overflow-align’ so the page number remains aligned with the outer edge of the text block:

```
<xsl:template name="Odd-Header">
  <fo:block
    keep-together.within-line="always"
    text-align="center"
    font-size="8pt"
    border-bottom="1pt solid black"
    axf:leader-expansion="force"
    padding-bottom="5pt"
    margin-bottom="4pt"
    axf:overflow-align="end">
    <fo:page-number color="transparent"/>
    <fo:leader />
    <fo:inline letter-spacing="0.22em">
      <fo:retrieve-marker
        retrieve-class-name="Chapter-Title"
        retrieve-position="last-starting-within-page" />
    </fo:inline>
    <fo:leader />
    <fo:page-number />
  </fo:block>
</xsl:template>
```

THE HONOR AND GLORY OF WHALING. 405
brawny doer of rejoicing good deeds, was swallowed down

Conclusion

Developing a stylesheet to format the first American edition of Moby-Dick by Herman Melville presented several challenges, but these were able to be solved by using a combination of the features of XSLT, XSL-FO, and AH Formatter extensions.

References

1. **IU Digital Library Program.** Moby-Dick, or, The Whale. Melville, Herman, (1819–1891). [Online]
<http://web.archive.org/web/20200815114219/http://webapp1.dlib.indiana.edu/TEIgeneral/view?docId=wright/VAC7237&brand=wright>.
2. —. Moby Dick, or, The Whale. [Online]
<http://web.archive.org/web/20201027153917/http://dogwood.dlib.indiana.edu:8080/xsubmit/rest/repository/wright/VAC7237.xml>.
3. —. Wright American Fiction. *Indiana University*. [Online]
<http://web.archive.org/web/20200725213048/http://webapp1.dlib.indiana.edu/TEIgeneral/welcome.do?brand=wright>.

4. **Antenna House.** Automated Analysis. *AH Formatter V7.0*. [Online] <https://www.antenna.co.jp/AHF/help/v70e/ahf-analyzer.html>.
5. —. Antenna House Formatter V7. [Online] <https://www.antennahouse.com/formatter-v7>.
6. **Grießhammer, Frank.** Introducing Source Serif 2.0. *Adobe Typekit Blog*. [Online] January 10, 2017. <https://blog.typekit.com/2017/01/10/introducing-source-serif-2-0/>.
7. **Wikipedia.** Moby-Dick. *Wikipedia*. [Online] https://en.wikipedia.org/wiki/Moby-Dick#Chapter_structure.
8. **Text Encoding Initiative.** Headers, Footers, and Similar Matter. *P5: Guidelines for Electronic Text Encoding and Interchange*. [Online] <https://tei-c.org/release/doc/tei-p5-doc/en/html/PH.html#PHSK>.