Layered Readings

Towards an electronic edition of Gabriel Harvey's Marginalia

Pilot Project Report

Introduction

The pilot project for the *Layered Readings* project was undertaken during the summer of 2004. The aim was to identify the requirements for an electronic edition of one of Gabriel Harvey's annotated books and a methodology that would meet those requirements. A paper based on the outcome of the pilot project was presented at the Digital Resources for the Humanities conference in September 2004.[1] This report is an extended and modified version of that paper.

The Limitations of the Printed Page

In the early twentieth century *Gabriel Harvey's Marginalia*, collected and edited by G.C. Moore Smith, were published in a limited edition.[2] The costs of printing and the small readership for such a work meant that Moore Smith was restricted in the amount of context that he was able to supply for the marginalia. It was possible to provide only the page reference and a brief indication of where on the page the note was found. This meant that the reader wishing to understand the textual context of the marginalia had to acquire a copy of the same edition as Harvey read. An understanding of the physical relationship of the note to the printed text required access to Harvey's own copy. Despite advances in the technology of printing in the last century, a modern editor of Harvey's marginalia would suffer from similar restrictions, if they wished to produce a printed book. It might be possible to include more contextual information, perhaps even images of the annotated pages. However, since Harvey annotated so many books – and we know of far more than Moore Smith did – it is unlikely that a publisher would provide as much context and as many images as the academic researcher would desire. Of course, within the academic community today we have comparatively easy access to electronic images of editions of early English books through *Early English Books On-line*[3]. So, the work of the researcher in marrying annotation to context is made that much easier. However, we believe an edition that married the marginalia with the context would provide an even richer research tool.

---

[1] DRH2004, Newcastle University, 5th-8th September, 2004.

[2] G.C. Moore Smith, *Gabriel Harvey's Marginalia* (Stratford-upon-Avon, 1913).

[3] http://eebo.chadwyck.com/home - a subscription is required to use this service.

Example 1: An annotated title page

Figure 1 shows the title page of a volume catalogued by Virginia Stern's *Gabriel Harvey*[4] as follows:

> [Hurault, Michel]. *An Excellent Discourse upon the Now Present Estate of France. Faithfully translated out of the French, by E.A.* [Edward Aggas]. London, John Wolfe, 1592.
> 4° STC 14005 Huntington 49490
> 'gabrielharvey. 1592'. 'given mee bie Mr Woolfe, for a special rare Discourse'. A long folded sheet pasted in between fols. 3 and 4 contains lists in Harvey's hand of: (1) duchies, with their dependent counties; (2) counties dependent on the crown; (3) peers of France; (4) archbishoprics; (5) academies. On reverse of the sheet in Harvey's hand: 'A compendious description of France. A proffitable Table'. On the final leaf: 'Gabrielharvey: this August: 1592. Il legere nutrica lo ingegno'.
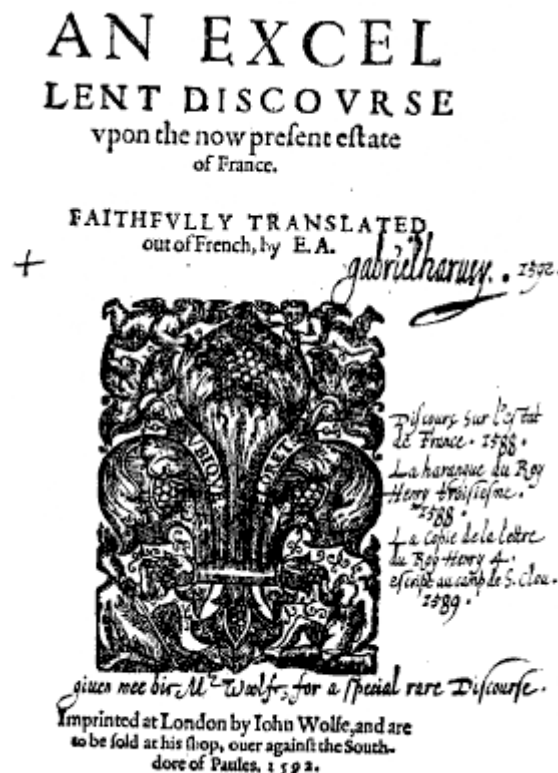

Figure 1

As we can see there is more annotation than is mentioned by Stern on the title page of this volume. The context of the annotation is clearly important. The signature indicates ownership and the annotation further indicates that the volume was given to Harvey by the printer at the time of publication. The notes to the right of the image apparently provide an aide memoire of the volume's contents. The cross to the left of the page below the title was part of the shorthand system with which Harvey marked his books.

---

[4] Virginia F. Stern *Gabriel Harvey: His Life, Marginalia and Library* (Oxford, 1979).

Example 2: A note linked to textual content

One of Harvey's annotated books was '*The Breuiary of Britayne. Writen in Latin by Humfrey Lhuyd .. Englished by Thomas Twyne. Imp. at London, by Richard Johnes. 1573.*' This book was annotated on the title page and flyleaf, at the end of the opening epistle and at the end of the text (i.e. on 3 printed and one blank page of a book with 96 folios plus extensive introductory matter). The annotation at the end of the epistle is given by Moore Smith thus:

Aiiij^v   [At end of '*Epistle*'-]
        Nihil turpius quam domi esse peregrinum: nihil magis pudendum, quàm ignaru*m* esse suæ Patriæ.[5]

It is unlikely that a facsimile edition of Harvey's marginalia would include images of all the pages that did not include annotations. However, in this case it is the previous page (Aiiii) that provides the context for Harvey's annotation:

> And so much the rather, for that in the studie of Geographie, it is expedient first to know exactly the situation of our owne home, where wee abide, before that wee shalbe able to judge how other countries doo lie unto vs, which are farre distant from vs, besides that it were a foule shame to be inquisitiue of the state of forreyne landes, and to be ignorant of our owne . As you honour beynge already perfectly instructed: is not now to learne at my hand.

An electronic edition needs to ensure that the context of all annotations is available to the user. This requires that the edition considers both the physical relationship of the annotation to the printed page and the logical links between the annotation and the printed text.

Defining the Problem

Our experience of working with projects that are encoding early modern manuscripts has taught us that they present interesting challenges. In the case of Harvey there is the added complication that the manuscript marginalia must be considered in conjunction with a printed text, while on occasion having no obvious connection to its content. The marginalia also represent the results of an often complex series of readings and annotations.[6]

---

[5] Nothing is more scandalous than to be a foreigner at home: nothing more disgraceful than to be ignorant of your native land.

[6] For more about Harvey's annotation practices, see Lisa Jardine & Anthony Grafton, "Studied for Action': How Gabriel Harvey read his Livy', *Past and Present* 129 (1990), 30-78.

Figure 2 presents a particularly complex example of Harvey's annotation, which demonstrates a number of common features.
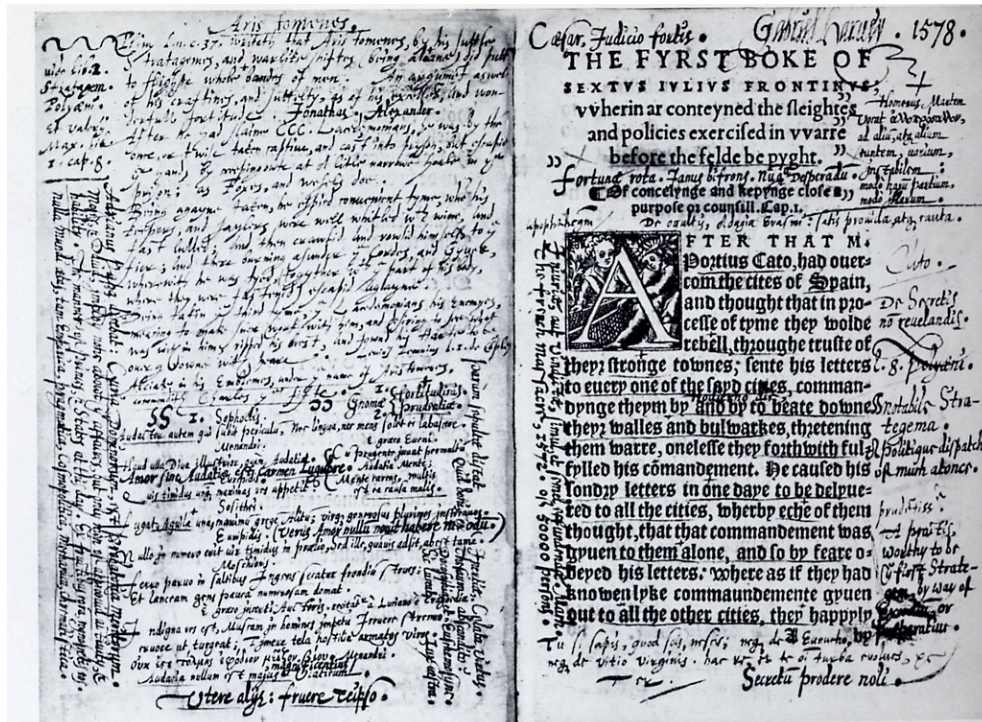

Figure 2

1. The formerly blank page on the left is covered in writing.
2. A date has been added in the top right corner of the right-hand page. These are common – both indicating dates of purchase and of successive readings of texts. From the title page of this volume we know that he purchased it in 1578 and added notes in 1580 and 1588.
3. He has also added a signature at the top of the right-hand page. This is also common – again it appears on the title page of this book, but the form of the signature is slightly different.
4. The notes occur in a number of languages. On this page we have English and Latin. Harvey also uses Greek, Italian, French and very occasionally Spanish.
5. He has underlined the printed text quite extensively. (In some books he made broad, coloured chalk-like markings the purpose and meaning of which are unknown.)
6. As well as writing in the margins and on the blank page, he has added notes between the lines of print.
7. The text is written in several directions. The direction of a note may be significant and should be recorded by the annotation.
8. Harvey made extensive use of symbols, such as those on the left hand page:

 and

These often provided a shorthand indication of the subject of a book or passage, but were not simply a shorthand reference – in his copy of Machiavelli's *Arte of Warre* Harvey decorated one battle scene with a Mars symbol at the end of every line, presumably for emphasis.

Identifying the Requirements

The following requirements were identified as essential for an electronic edition of one of Harvey's annotated books:

1. to record the underlying printed text in full.

2. to map the physical relationship of the annotations to the printed page.

3. to record the content, direction, style etc. of all marginalia;

4. to record all annotations of the printed text, such as underlining.

5. to provide the means for investigating the evolution of the text. (We want to try to recover not just how the text looks now, but how it looked when Harvey first bought it, when he re-read it ten years later etc. Obviously, this is a complex task – but we want the barriers to its achievement to be intellectual, not technical.)

6. to provide the reader of our eventual edition with multiple views of the marked up transcription: the physical page; the whole printed text of a chapter; the Latin annotations of a chapter; the printed text and the annotations – including underlining etc. - as they appeared on successive readings

7. to enable scholars to make various selections for analysis: all annotations in Latin or Greek or made in a particular year or from a particular source; all underlined text; all occurrences of a particular symbol or all symbols appearing in a particular position.

As we analysed what it was we wanted to do, it became obvious that a straightforward, single hierarchy approach to the encoding would not be sufficient. In particular we would like to be able to create different groupings of features as the occasion demands, rather than permanently assigning particular features to separate hierarchies.

Approaches Available within the TEI Guidelines

There are, of course, a number of approaches that can be adopted within the TEI Guidelines for dealing with complex hierarchies within a text.[7] Some of these we have direct experience of and for others we examined the experience of other projects.

Concur

Concur is an optional feature of SGML, which allows multiple hierarchies to be marked up concurrently in the same document. Since it is optional in SGML and not available in XML, there is limited support for this feature. The tools that CELL has developed for use in other projects depend upon XML and could not be used with concur. We would prefer not to have to define binding logical hierarchies within the encoding, since in different circumstances different hierarchies would serve our research purposes.

---

[7] The TEI P4 Guidelines are available on-line at http://www.tei-c.org/P4X. Chapter 31 discusses multiple hierarchies. It is due to be 'substantially revised' at the next release of the Guidelines, but the timetable for that release is currently unclear.

Milestone elements

Milestone elements are empty elements which mark the boundaries between elements in a non-nesting structure. It assumes the existence of independent, overlapping hierarchies. We are currently producing a new edition of manuscripts that were encoded around a logical hierarchy based on the intellectual content with the physical features such as page breaks marked as <milestone> elements. This has caused us headaches, because the new edition includes page images and we needed to match each physical page of text to a page image and have a co-coordinated scroll of the text and image. Working on the problem of implementing this functionality with the XML we had available helped to clarify our thoughts concerning the proposed Harvey edition. Our conclusion was that the use of milestone elements would require an unacceptably high level of programming support to fulfill our requirements.

Fragmentation

Fragmentation divides what is logically a single element into fragments, which nest satisfactorily within its context. This approach multiplies the number of elements, complicating mark-up and the analysis of content. It also requires the creation of separate logical hierarchies within the mark-up.

Virtual joins

Virtual joins allow for the recreation of virtual elements from fragments of text, which may be discontiguous or out of order. It is complicated to encode and assumes separate logical hierarchies.

Multiple Encoding

One solution would be to produce multiple encodings of the information, so that, for example, one set of mark-up related to the appearance of the physical pages, another to the printed textual content and another to the marginalia. This would be cumbersome, create redundant information and introduce the possibility of inconsistencies being introduced between the different sets of mark-up. It also assumes that the same logical divisions would fulfil all our research requirements, which is not the case.

Standoff Mark-up

In standoff mark-up the mark-up data is held separately from the textual content of the source.[8] To achieve our requirements more than one set of markup data would be required, which would necessitate the identification of separate logical hierarchies. Previous experience makes us wary of adopting this approach, since inconsistencies can occur when more than one physical file is involved in holding the markup. Problems can arise with the well-formedness of standoff markup, once it is incorporated with the source.

All these available methods had a fundamental disadvantage for us, as they rely on separating features into distinct hierarchies. From our experience with previous projects, we prefer to avoid multiple source files. We also prefer avoid placing heavy

---

[8] For those with long memories this is reminiscent of the approach adopted by Southampton University's *Microcosm* system of the 1990s, which held text in RTF files and links in a separate linkbase.

demands for consistency on encoders, as machines are better at managing consistency than humans.

The Proposed Solution

The approach we have adopted is to allow concurrent, overlapping hierarchies in the source file produced by the user. This solution was suggested by our experience within the commercial world of the use of the #ifdef / #ifndef feature of the C programming language to allow the multiple logical versions of source code required to support different products to be held within a single physical file.

For the pilot project the source file we created was processed to produce valid XML using fragmentation. The advantage of this approach is that the hard work is done by the Python program that creates the XML, not by the encoder. It also means that we could produce other parsers to use multiple encoding, virtual joins etc., if the type of processing required suggests this is the best approach. Different filters create different logical hierarchies from the same source by including or excluding tags. The treatment of elements – for example whether excluding a tag excludes its content – was defined within the parser for the pilot project, where only a few tags were involved. In an actual project we would institute a formal grammar for defining this, in conjunction with a tool for its creation and editing. The same tool would be used to maintain the XML schemas for the project.

Underlining
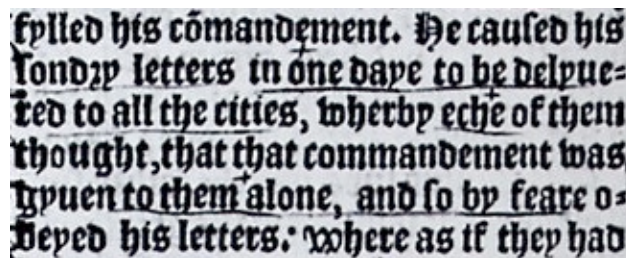
This is an extract from the page we looked at before. The user is required to tag what they see. So <underline> tags indicate the start and end of the underlining, ignoring any intervening markup.

```
<l>fylled his comandement. He caused his</l>
<l><underline revision='1'>sondry letters in one daye to be delyue=</l>
<l>red to all the cities</underline>, wherby <underline revision='1' >eche</underline>
of them</l>
<l>thought, that that commandement was</l>
<l>gyuen to <underline revision='1' >them alone and so by feare</underline> o=</l>
<l>beyed his letters: where as if they had</l>
```

The revision attribute enables us to produce outputs of the base text and the first level of annotation by including or excluding tags with a 'revision' of '1'.

```
…
<l frag='34'>fylled his comandement. He caused his</l>
<l frag='35'><underline revision='1' frag='36'>sondry letters in one daye to be
delyue=</underline></l>
<underline revision='1' frag='36'><l frag='37'>red to all the cities</l></underline>
```

HVY/2004/05/001. <http://www.livesandletters.ac.uk/>. AHRB Centre for Editing
Lives and Letters. Dr Jan Broadway 'Layered readings: Towards an electronic edition
of Gabriel Harvey's marginalia. Pilot Project Report'

```
<l frag='37'>, wherby <underline revision="1" frag='38'>eche</underline> of them</l>

<l frag='39'>thought, that that commandement was</l>

<l frag='40'>gyuen to <underline revision='1' frag='41'>them alone, and so by feare</underline>
o=</l>

<l frag='42'>beyed his letters: where as if they had</l>

…
```

Here the parser uses fragmentation to create valid XML from the input, inserting the
additional tags required and keeping track of related tags through the 'frag' attribute.
You will note that for the <underline> tags associated with frag 36, the revision
attribute is repeated. This is invisible to the user.

The parser has been written in Python and relies on regular expressions. For the pilot
project 3 basic parses were created: all tags, excluding requested tags and including
only requested tags. The parser would either return the tags and text or only the tags
requested (the latter being used, for example, to create input for client-side image
maps from <div> tags with x,y co-ordinates defined).

A further function is provided that uses the Python XML DOM module to convert the
XML to HTML.

```
…

fylled his comandement. He caused his<br />

<span class='underline'>sondry letters in one daye to be delyue=</span><br />

<span class='underline'>red to all the cities</span>, wherby <span class='underline'>eche</span>
 of them<br />

thought, that that commandement was<br />

gyuen to <span class='underline'>them alone, and so by feare</span>
 o=<br />

beyed his letters: where as if they had<br />

…
```

As the parser produces valid XML, XSLT could also be used for the transformation.
We also explored the design of an editor to facilitate encoding, but implementation of
this was outside the scope of the pilot project.

Textual Divisions

The text in the example used in the pilot project consisted of an underlying printed
text with annotations in all four margins running in different directions. The printed
text and each annotation is contained within a <div> element. The position of the
<div> on the physical page is given as the x,y co-ordinates of the top left hand corner
of the text within the division and of the bottom right corner according to the
direction in which the text is written. All the languages which Harvey used are
written from left to right and from top to bottom, so an analysis of the four co-
ordinates provides the information needed to understand the relationship of the text to
the physical page. If Harvey had used Hebrew, Japanese or other non-Western
languages, the interpretation of the co-ordinates would be dependent upon the
language content of the division. This would require additional intelligence to be built
into the analysis tools used by the project, but would not present a major difficulty.

Textual divisions such as paragraphs frequently cross the boundaries of physical
pages, so the multiple hierarchy approach discussed above is utilised to map the

HVY/2004/05/001. <http://www.livesandletters.ac.uk/>. AHRB Centre for Editing
Lives and Letters. Dr Jan Broadway 'Layered readings: Towards an electronic edition
of Gabriel Harvey's marginalia. Pilot Project Report'

relationship of logical textual components to physical pages. In the example below, each <div> is a child of a <page> element, so that the physical co-ordinates of the division can be mapped onto a particular page. The <p> elements, representing the logical divisions of the printed text into paragraphs, may cross the boundaries of <page> and <div> elements, as shown:

```
<text>
  <page num="1">
  <div rev='1' x1='541' y1='13' x2='734' y2='52' lang='latin'>
     <p>…</p>
  </div>
  <div x1='578' y1='46' x2='914' y2='173'>
     <p>…</p>
  </div>
  <div x1='625' y1='185' x2='861' y2='221'>
     <p>…</p>
  </div>
  <div x1='579' y1='234' x2='907' y2='649'>
     <p>
  </div>
  </page>
  <page num="2">
  <div x1='578' y1='46' x2='907' y2='649'>
     </p>
  </div>
  </page>
</text>
```

Conclusion

The challenges identified by this project have a far wider applicability than simply the encoding of Gabriel Harvey's marginalia. In *Radiant Textuality* Jerome McGann wrote:

> SGML processors have a poor aptitude for mark-up of textual features that are concurrent but logically distinct. A simple instance would be trying to permit a simultaneous mark-up of a book of poems by page unit and by poem. In SGML you are led to choose one or the other as the logical basis of the mark-up design.[9]

Over the past year CELL staff have had numerous discussions with researchers interested in manuscripts that present similar challenges to encoding using the established TEI Guidelines.

In our experience imposing strict hierarchies at the encoding stage results in marked-up texts which are less flexible than our research objectives require. Human researchers are good at perceiving and understanding the complexity of a source. Computer programs are good at consistently applying rules. We believe that the approach outlined above could be developed to enable men and machines to each do what they do best. We hope to take this research further over the next year and would welcome input from the research community.

Jan Broadway

---

[9] Jerome McGann, *Radiant Textuality* (2001), 90.