Bicentennial Bits and Bytes

The Pittsburgh Digital Frankenstein Project

Rikk Mulligan | Elisa Beshero-Bondar | Matt Lavin | Jon Klancher
@CritRikk | @epyllia | @mjlavin80 | @jklancher

MLA 2018: Saturday Jan. 6 @ 3:30pm; Sheraton Riverside Suite

Link to these slides: http://bit.ly/BicFrankMLA18
Hello. Thank you for sharing your time with us, especially late on a Saturday afternoon. Just to make sure you are in the right place, this is session 632: Bits and Bytes: The Pittsburgh Digital Frankenstein Project.

I and my fellow panelists will be describing several aspects of our project over the next forty minutes or so to set up our roundtable discussion with you.

I'm Rikk Mulligan; my fellow panelists are Elisa Beshero-Bondar, Matthew Lavin, and Jon Klancher. I'll begin by introducing our project and explaining how we came together, as well as what each of us are bringing to the project and how this helped us define our initial project plans. My co-panelists will go in depth into our current phase and we'll end with some discussion of our next steps and May 2018 milestone.
A Patchwork Team

- **Elisa Beshero-Bondar**, Director, Center for the Digital Text, University of Pittsburgh at Greensburg
- **Jon Klancher**, English Department, Carnegie Mellon University
- **Matt Lavin**, Director, Digital Media Lab, University of Pittsburgh
- **Raff Viglianti**, Maryland Institute for Technology in the Humanities (MITH), University of Maryland
- **Scott Weingart**, Program Director of Digital Humanities, University Libraries, Carnegie Mellon University
Most Humanities research is done by individuals, but Digital Scholarship tends to require a team, often directed by a principal investigator with a well-defined research agenda. Our team is somewhat different—we formed around the opportunity to contribute to a project focused on the bicentennial anniversary of the publication of Mary Shelley’s novel Frankenstein. We came together organically through online and face-to-face relationships because of the possibilities of working together rather than as part of a specific goal or design.

The Pittsburgh Digital Frankenstein Project began to coalesce during October 2016, but its beginnings go back to August 2016 when Neil Fraistat of MITH—the Maryland Institute for Technology in the Humanities—contacted Scott Weingart to ask if he might be interested in doing some visualizations on the materials in the Shelley Godwin Archive.

A couple of weeks later, while Scott and I were discussing an unrelated science fiction project, he asked if I had any interest in Frankenstein. I’d just learned that CMU has a copy of the first edition in our special collections and thought this would be a great opportunity to explore new DH methods and highlight our special collections through a DH project.

Scott then video chatted with Neil, Raff Viglianti, David Rettenmaier, and Purdom Lindblad at MITH in September to begin scoping out the project. He then sent out an email to the four of us in the Pittsburgh area (Elisa, Matt, Jon, and myself), introducing us and asking if we might explore potential goals.

Although Raff was not involved in our first meetings, Elisa was already in contact with him and had previously arranged for him to speak at Pitt-Oakland and Pitt-Greensburg on music encoding and its applications. Elisa and Raff had also been in touch with Neil about contributing an update to the edition of Frankenstein on the Romantic Circles website, one that would interweave with the Shelley-Godwin Archive’s edition of the Frankenstein manuscript Notebooks. Because Elisa and Raff work together as members of the TEI Technical Council, Neil hoped they might work together on the interconnection of those editions, as part of the “Pittsburgh group” collaboration.

Jon accepted Scott’s invitation because of his involvement in the Romantic Bicentennials as Co-Director of Networked Events, which has several upcoming Frankenstein projects. As the only Romanticist at CMU we were hoping he would participate and he was looking at this as a chance to work with a DH team.

Scott and I are members of CMU’s new digital center, dSHARP, and Matt is one of those who attended our open office/open consulting hours last year. After hearing about the project after our first couple of meetings he opted to join in the late fall of 2016.
What We Contribute

- **Elisa Beshero-Bondar**: *Romanticist*; Textual Scholar; TEI architecture and collation
- **Jon Klancher**: *Romanticist*; Book Historian; Annotations
- **Matt Lavin**: 19th Century Americanist; Textual Analysis, Stylometry
- **Rikk Mulligan**: 20th Century Americanist; Web Coding, Interface Design
- **Raffaele Viglianti**: Research Programmer; Shelley-Godwin Archive encoding; TEI pointers to S-GA Notebooks
- **Scott Weingart**: Early Modernist, *History of Science*; Textual Analysis, Stylometry

Our team is eclectic—a patchwork of institutions, roles, and levels of technical expertise and DH experience. This is our strength. Even as Victor’s “patchwork” construct is greater than the sum of its parts (especially in modern reimaginings), so is our goal to construct a digital edition with features that bridge print and digital resources. We also approach the project from very different perspectives and can learn a lot from each other. As Romanticists Jon and Elisa are our subject matter experts on Shelley’s era and the text; both have several years experience teaching a variety of Frankenstein editions. Scott expertise in the history and philosophy of science will also contribute to Jon’s annotations of the text. Elisa’s extensive experience with the Text Encoding Initiative technologies including TEI XML and XSLT transformations are providing the framework for our online texts. She has worked with Raff to architect our collations and consulted with David Birnbaum at the University of Pittsburgh on some of the thornier coding issues. Raff is also contributing from MITH to help us integrate the Shelley-Godwin Archive tags and annotations, as well as later helping with the TEI pointers for our future interface efforts.

Both Matt and Scott are DH generalists with extensive experience with a range of tools, methods, and projects. They began working on the textual analysis of our materials once Elisa and Raff got them in the shape they required. They are currently working on the stylometrics. [In this case, programmatic approaches to the study of measurable features of (literary) style, such as sentence length, vocabulary richness and various frequencies (of words, word lengths, word forms, etc.) with practical applications in authorship attribution research.] I only had a brief flirtation with TEI XML before this project. I’ve been working closely with Elisa to create the clean text files of the print editions and to integrate the notebooks into our corpus. My expertise lies more in web and interface design, which will come more into play as we evolve the presentation of our results.
Print publications

known/authorized by MWS

- 1818 Edition (3 volumes)
- 1823 Edition (2 volumes)
- 1831 Edition (1/2 of a volume)
  - bound with Friedrich von Schiller's *The Ghost Seer* in Bentley's Standard Series of novels

But I’m getting ahead of myself. Before we could decide what directions we might go in for our digital project, we needed to survey the current Frankenscape. Although Elisa, Jon, and Raff were well acquainted with Frankenstein, the rest of us were only passing familiar with an edition of the novel. We began by looking at the three print editions to see what we might contribute to the digital scholarship.

It is estimated that 500 copies of Frankenstein were published anonymously on January 1, 1818, in three small volumes. The current Romantic Circles edition attributes the Preface to Percy Shelley although neither the dedication or preface are signed or initialed in the actual copy. As a number of scholars have pointed out (Charles Robinson, Susan Tyler Hitchcock) the novel quickly inspired a number of stage adaptations. These proved so successful that Mary’s father, William Godwin, supervised the editing and republication of a two volume edition in 1823. The changes in this edition are so minor, which may explain why we could find no digital edition and had to digitize it ourselves from a photo facsimile. The last edition released during Mary Shelley’s lifetime appeared in 1831. It is important to note that the changes from the 1818 and 1823 to the 1831 are fairly extensive, and that everyone who involved in the ghost story contest and earlier text other than Mary had died by the time she released this edition with her story of its genesis in the introduction.
Digital Sources

- **Pennsylvania Electronic Edition**
  - 1994 start (early HTML, frame-based)
  - [http://knarf.english.upenn.edu/](http://knarf.english.upenn.edu/)

- **Romantic Circles**
  - 2009 update from HTML to TEI XML
  - [https://www.rc.umd.edu/editions/frankenstein](https://www.rc.umd.edu/editions/frankenstein)

- **Manuscript Notebooks**
  Bodleian Arbinger c56, c57, c58
  - Shelley-Godwin Archive
  - TEI XML
  - [http://shelleygodwinarchive.org/contents/frankenstein/](http://shelleygodwinarchive.org/contents/frankenstein/)

We started with a goal to prepare an updated and improved digital edition of Mary Shelley's Frankenstein that conforms to the TEI P5 standard. Elisa and Raff knew that much of this material already existed online.

The earliest work dates back to the 1990s, in Stuart Curran and Jack Lynch’s University of Pennsylvania Electronic Edition. This website uses frames to display the 1818 and 1831 editions, as well as a “variorum” frame to show differences. This site also hosts hundreds of additional files as hypertext annotations and connections.

The Romantic Circles website is a refereed scholarly resource; it published versions of the 1818 and 1831 editions upcoded from the PAEE HTML to TEI 3 in 2009. This site also offers a visual comparison of the texts, hosted by the Juxta Commons.

Finally, the Shelley-Godwin Archive has the transcribed Abinger Manuscript Notebooks, c56, c57 and c58 currently in the Bodleian Library of Oxford. The archive focuses on the manuscripts, providing access to the extant fair copy of Frankenstein in three forms and a range of critical materials. [(1) the physical order of Bodleian MS Abinger c.58; (2) the virtually reconstituted order of Notebooks C1and C2; and (3) the linear chapter sequence of the three-volume fair copy.]
1974  Rieger: inline collation of "Thomas" w/ 1818, 1831 variants in endnotes

~mid-1990s  


1996  

Crook crit. ed of 1818, variants of "Thomas", 1823, and 1831 in endnotes (P&C MWS collected works)

C. Robinson, *The Frankenstein Notebooks* (Garland): print facsimile of 1816 ms drafts

2007  

*Romantic Circles* TEI conversion of *PAEE*; separates the texts of 1818 and 1831; collation via Juxta

2013  

*Shelley-Godwin Archive* publishes diplomatic edition of 1816 ms drafts

2017  

*Pittsburgh Bicentennial Frankenstein Project* begins: assembly/proof-correcting of PAEE files; OCR/proof-correcting 1823; "bridge" TEI edition of S-GA notebook files; automated collation; incorporating "Thomas" copy text
To develop our “improved and updated” digital edition we needed to think ask the question: “what are the authoritative scholarly and critical editions of Frankenstein”?

We've gone back to looking at James Rieger’s University of Chicago Press annotated edition from 1974. It is one of the first to emphasize the 1818 text; it also included the “Thomas” edition of 1823 (Mary Shelley annotated 1818 copy) in its analysis.

The first online collated edition, Curran and Lynch had only compared the 1818 and 1831 editions.

Nora Crook and Charles Robinson made path-breaking scholarly editions--one critical, and one of the Abinger ms notebooks--in print, not connected to Curran's work.

Romantic Circles might be considered an update if not an upgrade of the PAEE, using programmatic tools in Juxta Commons rather than hand-collation.

The Shelley-Godwin Archive went online with the Notebooks digital edition in 2013, making a diplomatic edition publicly available.

Our work is a bridge: We bring all of these nodes of scholarship together by considering the Abinger manuscript notebooks, the 1818, 1823, and 1831 editions, and Mary Shelley’s annotations in the “Thomas” edition. We do not seek to replace the scholarly apparatus created by others, but to build upon, correct, update, and connect this fine work. We are cognizant that some of the work has errors we can fix or was built with almost-obsolete web technologies--needing a more stable digital architecture.
Evolving Project

Returning to the original texts to produce:

- Clean Text files for each edition (1818, 1823, 1831)
- TEI XML files for each edition
- Comprehensive collations from ms through 1831 to bridge and build on previous critical editions (print and digital)
- Variorum interface to show changes over time
- Stylometric analysis
- Annotations

https://github.com/ebeshero/Pittsburgh_Frankenstein
We have completed several steps toward preparing a new, scholarly digital edition of Frankenstein in TEI 5. We have also started the work to offer additional scholarly resources online as part of this project. We maintain a GitHub repository to track and share our work and render our efforts as transparent as possible. As of now we have completed plain text editions of the 1818, 1823, and 1831 editions on our GitHub. We have added comments about our initial attempts to use the PAEE and RC materials, and explained the process we used to produce the clean text files. Elisa will go into more detail regarding our efforts to prepare the TEI XML editions and collation files. She will also present our work to date on developing a comprehensive collation of all five sources, and sketch out our goals in developing a visual variorum interface to display the differences between them.

Matt will discuss our current work on the stylometric analysis. For those who are unfamiliar, stylometrics in this cases uses programmatic approaches to study the measurable features of (literary) style; this can include sentence length, vocabulary richness and the various frequencies of words, word lengths, and word forms as part of authorship attribution research. Jon will talk to you about how we hope to augment the annotations of previous scholarly editions and both the online resources with something different.
Variorum project

- Manuscript:
  (Notebooks: Abinger c56, c57, c58)
- "Thomas copy" Edition (1818 edition with hand annotations by Mary Shelley)
- 1818 Edition (3 volumes)
- 1823 Edition (2 volumes)
- 1831 Edition (1/2 of a volume)

Illustrations by Bernie Wrightson.
Building a Digital Variorum

Elisa Beshero-Bondar
@epyllia
Motivating Questions

- Can we make an edition that conveniently compares the manuscripts to the print publications?

- Can we make a comprehensive collation to show changes to the novel over time, from 1816 to 1831?
  - How many versions? (5 and a bit?)
  - Which editorial interventions persist from 1816 to 1831?
    - MWS in the "Thomas" copy: how much of this persists into 1831?
    - PBS's additions: which/how many of these persist to 1831?
    - What parts of the novel were most mutable?
Our Project Genealogy: Critical and Diplomatic Editions Leading to the Pgh Frankenstein Project

1974
Rieger: inline collation of "Thomas" w/ 1818, 1831 variants in endnotes

~mid-1990s

1996
Crook crit. ed of 1818, variants of "Thomas", 1823, and 1831 in endnotes (P&C MWS collected works)

C. Robinson, *The Frankenstein Notebooks* (Garland): print facsimile of 1816 ms drafts

2007
Romantic Circles TEI conversion of *PAEE*; separates the texts of 1818 and 1831; collation via Juxta

2013
Shelley-Godwin Archive publishes diplomatic edition of 1816 ms drafts

2017
Pittsburgh Bicentennial Frankenstein Project begins:
assembly/proof-correcting of PAEE files; OCR/proof-correcting 1823; "bridge" TEI edition of S-GA notebook files; automated collation; incorporating "Thomas" copy text
"We stand on the shoulders of giants." As Rikk mentioned, we're building on a lineage of Frankenstein edition work that illuminates the novel's gestation and transformation.

Rieger: INLINE collation: wants to show MWS's process, as she handwrote alterations on a copy of the 1818 Frankenstein. As you read Rieger, you see a system of markup distinguishing MWS's hand from the print, showing her insertions and additions. It's important for FOREGROUNDING the act of comparing texts in the line of the reader's sight--precedent for Curran's PA Electronic Edition.

1990s: BURST of activity! Frankenstein was a major EARLY experiment to build a scholarly apparatus in hypertext in the early years of the World Wide Web. The web and print editions improve the scope of comparison--beyond just 1818 vs 1831. Charlie Robinson's facsimile edition of the Abinger manuscript notebooks illuminates the writing process and hands at work over the novel. Much more is now available to the scholar to document CHANGE OVER TIME.
The dream of the 90s. . .

Hypertext / Hypercard books and the *PAEE*

*Frankenstein; or, the Modern Prometheus*

By **Mary Wollstonecraft Shelley**

- Accessing (reading, writing, editing) texts in nonlinear ways
- Multiplying and individualizing points of access
The dream of the '90s: *Frankenstein's* inspiration for hypertext experiment

- Roughly contemporary with the PAEE + mid '90s scholarly edition efforts
- What if the female creature survived and had a chance to create her own story with lots of options?
- Experimental nonlinear navigation...hundreds of hypercards...plot your own course
I've included this slide to emphasize how the 1990s marked a moment of experiments with Frankenstein as a body of text--The building of a Creature and the assemblage of a textual body become continuous activities--the reader participates. It isn't just that we theorized with Stanley Fish about readers assembling texts for themselves; the experimenters with Frankenstein like Shelley Jackson and Stuart Curran invited readers to engage actively in remixing, juxtaposing, and exploring options.
hundreds of small html files, juxtaposed in frames
Discuss HOW this works as collation and what's so exciting about it.
The dream of the '90s is alive…
(in Pittsburgh)

Digital Collation for a "Variorum" interface

- select a text from what version the reader chooses:
  - 1816 MS | 1818 | "Thomas" | 1823 | 1831
- compare that text to what version the reader chooses
- view the "molten" portions of the novel in context with the stable portions
- navigate multiple texts in context with one another
- make the critical apparatus a vantage point: see how the novel changed over time without having to find the fine-print endnotes

source: I programmer article on "Frankenstein" malware

The Creature of Collation?
We make newly formed text "bodies" from disparately formed source materials.
TEI: The Text Encoding Initiative

- a community-maintained standard
- 1987 @ Vassar: draft of Poughkeepsie Principles
  "provide a standard format for data interchange in humanities research.
- Guidelines for the Encoding and Interchange of Machine Readable Texts: first drafted 1990; published on the web by 1999 (P3)
- Standards for encoding texts co-evolve with standards for developing human and machine-readable markup languages
  - HTML (w3c) || (early) SGML and XML
- TEI XML tree structure:
  - meant to store a stable format not subject to commercial processing requirements
- possible to publish TEI directly or convert to HTML; PDF; TEX; other document formats.
SGML = standard generalized markup language

XML = eXtensible markup language
From PAEE to Pgh Variorum...

values in common

1. Small pieces are optimal for collation.
2. There is no single "complete" edition.
3. Each output (plain text, XML, TEI collation) = **viable edition on its own**.
4. Interface invites the user to play: put the pieces together.

image source: a friend's Lego set
A Bridge-Building Challenge

- Reconcile multiple kinds of text encoding:
  - old '90s HTML (1818, 1831)
  - not-so-plain OCR-generated text (1823)
  - TEI XML for manuscripts: (S-GA diplomatic edition)

- Construct "Bridge" XML for collation
  - Markup-assisted machine collation (collateX):
  - "flattened" XML hierarchies for even collation units
  - ms metadata markup (e.g. "hands") to ignore in collation, but preserve in the output
  - pointers outward to manuscript editions (S-GA, Morgan Library)

Pittsburgh's bridges (1963)
Source: NewsCastic.com
Collation "stitchery"

- Can be done by hand in TEI
- Automated: via CollateX
  - Algorithms for locating union and "delta" points in "streams" of text
  - Inputs in a variety of formats (XML/TEI, plain text, JSON)

- Output / Visualization options:
  - Text table (above); SVG flow chart; XML
  - JuxtaCommons on the web
  - Develop a custom web interface (via XML output)
<surface><graphic>
</graphic>

I<hi rend="smallcaps">T</hi>

It was on a dreary night of November that I beheld the accomplishment of my toils. With an anxiety that

...
<table>
<thead>
<tr>
<th>CH.</th>
<th>CH.</th>
<th>CH.</th>
</tr>
</thead>
<tbody>
<tr>
<td>—</td>
<td>IV.</td>
<td>IV.</td>
</tr>
<tr>
<td>&lt;surface&gt; &lt;graphic&gt;</td>
<td>&lt;/head&gt; &lt;p&gt;</td>
<td>&lt;/head&gt; &lt;p&gt;</td>
</tr>
<tr>
<td>&lt;/graphic&gt;&lt;lb/&gt;&lt;mile&lt;br&gt;stone spanTo=&quot;#c56-0045.04&quot; unit=&quot;tei:head&quot;&gt;Chapter 7&lt;hi rend=&quot;smallcaps&quot;&gt;T&lt;/hi&gt;</td>
<td>I&lt;hi rend=&quot;smallcaps&quot;&gt;T&lt;/hi&gt;</td>
<td>I&lt;hi rend=&quot;smallcaps&quot;&gt;T&lt;/hi&gt;</td>
</tr>
<tr>
<td>was on a dreary night of</td>
<td>was on a dreary night of</td>
<td>was on a dreary night of</td>
</tr>
<tr>
<td>November</td>
<td>November,</td>
<td>November,</td>
</tr>
<tr>
<td>&lt;del&gt;that I beheld &lt;del rend=&quot;strikethrough&quot;&gt;&lt;add hand=&quot;#pbs&quot; place=&quot;superscript&quot;&gt;the frame on which&lt;/add&gt;&lt;/del&gt; &lt;/lb/&gt;</td>
<td>that I beheld the accomplishment of</td>
<td>that I beheld the accomplishment of</td>
</tr>
<tr>
<td>my</td>
<td>my</td>
<td>my</td>
</tr>
<tr>
<td>man completed &lt;del add place=&quot;#intra linear&quot;&gt;&lt;/add&gt;&lt;add hand=&quot;#pbs&quot; place=&quot;intra linear&quot;&gt;&lt;/add&gt;&lt;del rend=&quot;strike through&quot;&gt;And&lt;/del&gt;&lt;lb/&gt; with an anxiety that almost amounted</td>
<td>toils. With</td>
<td>toils. With</td>
</tr>
<tr>
<td>to to to</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(image source: S-GA)
A running text stream...?  
Or an architecture of bridges?

(collateX SVG output)
XML collation: flagging variants and Percy's hand
Stylometry and Digital Frankenstein

Matthew Lavin
University of Pittsburgh
@mjlavin80
Research Questions

1. How does *Frankenstein* change stylistically across different expressions/manifestations?

2. How can those changes be attributed and/or characterized?
   - In whose authorial voice is *Frankenstein*? Direct analysis of Percy and Mary

3. Do stylistic changes affect how *Frankenstein* reads in relation to cultural categories like genre, “modernness,” linguistic register generally, and scientific vocabulary? If so, how?
Outline of Exploratory Measures

In notebooks, term counts/relative frequencies of:

- Mary’s hand initial
- Mary’s hand strikethrough
- Percy’s hand suggested vs. adopted
- Mary’s hand revised (sometimes Mary ver1, ver2, final, etc.)

Across our three print editions:

- Term counts/relative frequencies of each text
- Term frequencies weighted against frequency across all documents (TF-IDF)
## Term Counts

Absolute Values of Term Count Differences across Editions, 1818 to 1823 (left) and 1823 to 1831 (right)

<table>
<thead>
<tr>
<th>index</th>
<th>word</th>
<th>1818_1823</th>
<th>1823_1831</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5814</td>
<td>7</td>
<td>220</td>
</tr>
<tr>
<td>1</td>
<td>2662</td>
<td>7</td>
<td>169</td>
</tr>
<tr>
<td>2</td>
<td>352</td>
<td>6</td>
<td>142</td>
</tr>
<tr>
<td>3</td>
<td>7175</td>
<td>6</td>
<td>103</td>
</tr>
<tr>
<td>4</td>
<td>1365</td>
<td>6</td>
<td>98</td>
</tr>
<tr>
<td>5</td>
<td>6397</td>
<td>6</td>
<td>93</td>
</tr>
<tr>
<td>6</td>
<td>5404</td>
<td>5</td>
<td>75</td>
</tr>
<tr>
<td>7</td>
<td>711</td>
<td>5</td>
<td>69</td>
</tr>
<tr>
<td>8</td>
<td>5591</td>
<td>5</td>
<td>66</td>
</tr>
<tr>
<td>9</td>
<td>1311</td>
<td>5</td>
<td>42</td>
</tr>
<tr>
<td>10</td>
<td>3570</td>
<td>4</td>
<td>41</td>
</tr>
<tr>
<td>11</td>
<td>6767</td>
<td>4</td>
<td>37</td>
</tr>
<tr>
<td>12</td>
<td>3033</td>
<td>4</td>
<td>35</td>
</tr>
<tr>
<td>13</td>
<td>506</td>
<td>4</td>
<td>34</td>
</tr>
<tr>
<td>14</td>
<td>2604</td>
<td>4</td>
<td>33</td>
</tr>
<tr>
<td>15</td>
<td>6575</td>
<td>4</td>
<td>33</td>
</tr>
</tbody>
</table>
### Relative Term Frequencies

<table>
<thead>
<tr>
<th>index</th>
<th>word</th>
<th>1818_1823</th>
<th>1823_1831</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>shewed</td>
<td>0.001555</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>fulfill</td>
<td>0.001333</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>daemon</td>
<td>0.001333</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>showed</td>
<td>0.000920</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>shewn</td>
<td>0.000889</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>daemon</td>
<td>0.000790</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>fulfil</td>
<td>0.000789</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>an</td>
<td>0.000685</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>shown</td>
<td>0.000677</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>phænomena</td>
<td>0.000667</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>aye</td>
<td>0.000667</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>when</td>
<td>0.000613</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>but</td>
<td>0.000570</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>may</td>
<td>0.000512</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>avalanche</td>
<td>0.000508</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>phenomena</td>
<td>0.000507</td>
<td></td>
</tr>
</tbody>
</table>

### Absolute Value of Weighted Term Frequency Differences across Editions (tf-idf), 1818 to 1823 (left) and 1823 to 1831 (right)
Collational Alignment

Types of Changes:

- Spelling normalizations
- Punctuation
- Word insertions, substitutions, deletions
- Word to phrase or phrase to word
- Reordering
Chapt. 2
Those events which materially influence our future destinies are often caused by slight or trivial occurrences. Strange as the simple fact may appear my fate had been regulated by a desire for knowledge in my early years to state those facts which led to my love pursuit of that study. When I was eleven years old we all went on a party to Thonon and were confined there by the inclementy of the weather obliged us to remain a day confined to the inn. In this house I chanced...
Punctuation Matters ... but not for all measures
Dynamics of DH Collaboration

The workflows and analytical paradigms of “machine learning DH” and “scholarly editing DH” are not factory fitted to one another, but they can be adapted to work in tandem. The gains are more valuable than the cost of the retrofit.
Dynamics of DH Collaboration

How can a single, carefully curated edition or set of editions be worked into a “macroanalysis” model where many uncorrected, dirty OCR texts are being compared to one another?

What kinds of questions can we ask with hand-corrected editions that we cannot ask with HTRC corpora?
Open Data and Reproducibility

I have argued elsewhere that openness invites open discussion and collaboration. It doesn't guarantee that these things will happen, but closed data practices all but guarantee that these practices will be difficult or impossible.
Next Steps: Questions/Methods

How can we characterize changes by trends established in analysis of each person’s hand?

How can we think about changes as moving closer or further away from a genre baseline?

How do index quantifications like “how modern” or “how scientific” each version of the text is? How do we account for “modern” and “scientific” as rapidly changing ideas?
As Scott Weingart’s work has shown, the computational determination of “who wrote what” has become less central to DH inquiry. I believe strongly, however, that these approaches will have a second life as we turn increasingly toward the use of authorship attribution techniques to study authorship as an historical and social construction.
Source: Web Annotation Data Model
(w3c Recommendation of 23 Feb 2017)
Annotated Print Editions of Frankenstein


My father had a sister, whom he tenderly loved, and who had married early in life an Italian gentleman. Soon after her marriage, she had accompanied her husband into [his] native country, and for some years my father had very little communication with her. About the time I mentioned she died; and a few months afterwards he received a letter from her husband, acquainting him with his intention of marrying an Italian lady, and requesting my father to take charge of the infant Elizabeth, the only child of his deceased sister. "It is my wish," he said, "that you should consider her as your own daughter, and educate her thus. Her mother’s fortune is secured to her, the documents of which I will commit to your keeping. Reflect upon this proposition; and decide whether you would prefer educating your niece yourself to her being brought up by a stepmother."

My father did not hesitate, and immediately went to Italy, that he might accompany the little Elizabeth to her future home. In 1818 is corrected to “his” in 1823 (Ch. I, 44).

Percy’s sister and mother had the name “Elizabeth.” Mary layer several family names into her narrative. The father’s disposal of a child, here a seemingly incidental plot device, is related to the multiple versions of parental care and carelessness in the novel. It is striking to see a father cast off his only child within a few months of her losing her mother. Would the gentleman have discarded a son?

A “stepmother” invokes a fairy-tale lore (and its social basis) of coolness to a stepchild. Mary’s stepmother (the second Mrs. Godwin) was less than loving, and her father would send his daughter, at age thirteen, to live in Scotland.

In 1831, Mary Shelley revised Elizabeth’s relation from a close blood tie (first cousin) to an adopted orphan—sharpening the contrast of the beautiful orphan’s good fortune in finding a loving family and the misfortunes of the abject ugly Creature (even with a connection to the Franksteins).
was the eldest, and the destined successor to all his labours and utility. No creature could have more tender parents than mine. My improvement and health were their constant care, especially as I remained for several years their only child. But before I continue my narrative, I must record an incident which took place when I was four years of age.

My father had a sister, whom he tenderly loved, and who had married early in life an Italian gentleman. Soon after her marriage, she had accompanied her husband into her<sup>10</sup> native country, and for some years my father had very little communication with her. About the time I mentioned she died; and a few months afterwards he received a letter from her husband, acquainting him with his intention of marrying an Italian lady, and requesting my father to take charge of the infant Elizabeth,<sup>11</sup> the only child of his deceased sister. It is my wish," he said, "that you should consider her as your own daughter, and educate her thus. Her mother's fortune is secured to her, the documents of which I will commit to your keeping. Reflect upon this proposition; and decide whether you would prefer educating your niece yourself to her being brought up by a stepmother."<sup>12</sup>

My father did not hesitate, and immediately went to Italy, that he might accompany the little Elizabeth to her future home. I have often heard my mother say, that she was at that time the most beautiful child she had ever seen, and shewed signs even then of a gentle and affectionate disposition. These indications, and a desire to bind as closely as possible the ties of domestic love, determined my mother to consider Elizabeth as my future wife; a design which she never found reason to repent.

From this time Elizabeth Lavenza became my playfellow, and, as we grew older, my friend. She was docile and good tempered, yet gay and playful as a summer insect. Although

Published in 1791, in the wake of the French Revolution (Volney was part of the Revolutionary government), *Les Ruines; ou Meditation sur les revolutions des empires* appeared in English as *Ruins, or Meditations on the Revolutions of Empires*, in 1792.

More properly, *The Ruins, Or, Meditation on the Revolutions of Empires; and the Laws of Nature*, by Constantin-François Chasseboeuf, who took the name Volney, published in 1791 in French. It was translated in 1802 into English. The book is described by Frankenstein scholar Pamela Clemit as a “powerful Enlightenment critique of ancient and modern governments as tyrannical and supported by religious fraud” (“Frankenstein, Matilda, and the Legacies of Godwin and Wollstonecraft,” in *The Cambridge Companion to Mary Shelley*, ed. Esther Schor [Cambridge: Cambridge University Press, 2003] 35.)

In light of the date of translation, the book in question must have been the French edition, and Safie and the creature learned French....

Of the books the Creature hears read aloud in the forest, Volney's *The Ruins; or, A Survey of the Revolutions of Empires* (1792) was the most closely associated with Europe's radical Enlightenment. (It was first published in French as *Les Ruines: ou Meditation sur les revolutions des empires* in 1791.) The Creature learns an illuminating critique of imperialism and exploitation from Volney, even as he also absorbs some of the Enlightenment's own prejudices ("slothful Asiatics"). The effect on the Creature is to give him a sense of the social or structural and not only a personal framework for understanding virtue and suffering. On Volney's role in the novel, see also Ian Balfour, " Allegories of Origins: Frankenstein after the Enlightenment," *SEL: Studies in English Literature 1500-1900* 56.4 (2016): 777-98.
using the hypothes.is tool for digital annotation with tags
Annotations that Tunnel through the Texts
(not only pointing outside)

- travel/expedition: Walton
- domestic affection (Walton - Margaret Seville)
- law / judicial system (Justine)
- travel/expedition: Victor
- domestic affection (DeLaceys and Safie)
- travel/expedition: Clerval
- domestic affection (Frankenstein family)
- law / judicial system Felix DeLacey
- travel/expedition: Creature
- law / judicial system Victor/Kirwin
Annotations in the Variorum Interface

Source: 6Sqft: "See New York City's subway lines superimposed over an aerial photo of the city" (2015)
Annotations and Intertextuality
"tunneling" through the texts
distinct from external pointers to context
affected by collation "bridges"
goal: available in each reading view
id markers signal textual locations:
book | chapter | paragraph
collation alignment shows corresponding locations in other versions
Portable hypothes.is annotations: "pinnable" by text string and id position markers
Frankenstein's invitation/challenge:

- **Build Digitally:**
  - Experiment with human and machine reading

- **Build a Strong, Sustainable Bridge:**
  - Update Romantic Circles edition
  - Interlinks to Shelley-Godwin Archive
  - Notebooks: point to ms pages
  - Morgan Library "Thomas copy"

- **Centralize the Critical Apparatus**
  - A tool for scholars
  - A metanarrative?
  - A remixing of the **reading process** for all who care about *Frankenstein*
  - Make all the texts available to all the readers

---

The work continues...

- Collation
- Annotation
- Stylometry
- Visualization and Variorum Interface