Reading ~~the~~ Changes: Variance and Digital Tools

Ellen Barth

Textual variants are the differences found in one copy of a printed text to another. We all know the famous line: "To be or not to be, that is the question[,]" but few among us know the variant form: "To be, or not to be, I there's the point [.]" Both were printed and published as the work of Shakespeare, from the First Folio and First Quarto,[1] respectively, but only the first version is taught and performed, as it is thought to be closest to the author's original (Parsons 87). Out of variants like this grew the field of textual criticism, in which scholars compare copies of a single edition, often with the aim of uncovering the most authentic and authoritative version of the text to create critical editions (Tanselle 1). The Bible and the works of Shakespeare have traditionally been two common foci of the textual scholar's gaze.

But today, unless working in textual criticism, it is unlikely that anyone would notice or even think about variants in their printed reading material, and one may even question whether variants still occur. In fact, they do—sometimes in abundance. And when they occur, digital collation tools can assist in comprehensive textual comparisons and allow readers to see and read variants in ways that would have been impossible decades ago.

From Scribe to Press

In textual communication, there has always been variation from copy to copy and text to text. The scribe, when depicted visually, is often seen seated at a writing desk, quill in one hand, and knife in the other. If a mistake was made while copying, the scribe would use this knife to scrape away the top layer of parchment where the incorrect text was written, leaving it clean to be written on again (Biggs). The fact that scribes are so often depicted with this tool in hand gives some indication as to the frequency of its use.[2] In addition to accidental errors made when copying, scribes are known to have purposefully changed texts as well. Henry Notaker writes: "Many of the well-trained and educated *scriptores* in

---

[1] Anyone interested in exploring variance in Shakespeare can visit the British Library website, which offers an online comparison of Shakespeare Quartos: https://www.bl.uk/Treasures/SiqDiscovery/UI/search.aspx.

[2] The knife was also used to sharpen quills.

the Middle Ages consciously altered language and content. They skipped material they had no interest in or did not find worth copying and added comments that eventually became a part of the text" (30). Chaucer famously wrote a poem about this very scenario, warning his scribe, Adam, not to rewrite his texts anymore while copying them or be cursed to suffer from a scalp disease (Mize 352).

Printing offered a consistency in mass quantity that could not be achieved in scribal times, as it is "a process designed to make multiple copies of identical items" which is "both cheaper and more accurate than the work of a scribe copying a manuscript" (Feather 5). Identical copies were not always the reality, as the words of Hamlet show,[3] but there is no arguing that after the invention of the printing press, the number of variants present in texts decreased drastically. In the present day, readers have gained near total confidence in the accuracy and consistency of their reading materials. To quote Adrian Johns: "You may safely assume that the book you now hold will […] be identical in all relevant aspects to one bought in the United States or in Great Britain" (255). However, the recent case of the bestselling novel *Cloud Atlas* and its version variants tells a very different story.

21st Century Variants

Cloud Atlas, an elaborate novel consisting of six nested stories, was written by British author David Mitchell. The book was published simultaneously in both the US and the UK in 2004, and it quickly became a bestseller, winning the British Book Awards Literary Fiction prize and later adapted into a feature film. For a book published in the US and the UK, a normal practice is for it to receive what is called 'trans-editing,' that is, small changes made to vocabulary and grammar to match the target reading group. This could include changing the word 'favor' to 'favour,' or 'row' to 'argument.' In the case of *Cloud Atlas*, however, the US and UK versions have been found to contain an "astonishing degree" of variants that go far beyond trans-editing (Flood).

Literary scholar Martin Paul Eve first noticed the differences. Eve was doing a close read of the novel when he found that one section in particular contained a large number of differences. The chapter in question is "An Orison on Sonmi~451," in which the narrative is organized as a series of questions and answers. Eve found several startling differences. First, the amount of material in each version is different, with the UK version containing more questions than

---

[3] The variance between two printed editions is perhaps not terribly surprising; however, as Charlton Hinman notes: "For more than two centuries the commonly employed methods of press correction were such that different copies of the same edition of a given text could not fail to be variant" (281).

the US version. There are structural differences, meaning that plot points occur at different periods in the story depending on the version (Eve 3–5). There are also many textual differences, which Eve notes as being substantial enough to impact literary critique (22).

For example, the UK version says:

> *No other version of the truth has ever mattered to me.*

In comparison to the US version:

> *TRUTH IS SINGULAR. ITS 'VERSIONS' ARE MISTRUTHS.* (Eve 20)

The words and typography vary to such a degree as to change the tone and reading, if not the function of the text. Moreover, Eve notes that nearly every sentence in the "An Orison on Sonmi~451" chapter contains differences across the US and UK versions (11). In contrast to Johns' assertion about the reliability of transatlantic published material, Eve states that: "[R]eaders of *Cloud Atlas* based in the US are likely to encounter a novel that stands starkly apart from that bearing the same title in the UK" (2).

Unlike the authors of most works examined for textual variants, David Mitchell is still alive; so Eve went directly to the source. In response, Mitchell sta-

ted that the differences were due to a "combination of chance and […] inexperience" (23). The manuscript, his story goes, was passed to US and UK editors (the US version being briefly "orphaned" before being taken over by a new editor), and changes to one version were not always sent to the editors of the other. Being a relatively new author, Mitchell himself did not keep track of the changes he made or to whom he sent them (23–24). The result was two different *Cloud Atlas* manuscripts published in two different nations.

A fluke? Perhaps. But with *Cloud Atlas*, Eve showed that variants still exist in our modern reading materials.

Finding Variance

In the past, when looking to compare—or collate—texts, collation had to be done mechanically, such as with a machine like the Hinman collator. This machine was developed by Charlton Hinman for the purpose of collating Shakespeare's First Folio (Hinman 280). It uses mirrors and flashing lights, allowing the operator to see variants 'dance' on the page and thereby identify them. With the Hinman collator, comparison can only be done across a single imprint or pressing as the pages need to line up for visual comparison (281). For many textual scholars, this is adequate; however, digital technologies have expanded the possible scope of textual comparison.

The digital humanities have opened up new avenues for the intrepid researcher, and when it comes to textual collation, scholars can now choose from numerous collation software applications, such as *CollateX*[4], *Versioning Machine*, and *Juxta*, to name just a few. In each case, texts—or witnesses—are entered into the application and then digitally compared. There are still chances for human error when inputting the witnesses, and if texts were scanned before entry, the possibility for computer recognition error remains as well. Nevertheless, there are clear advantages to digital collation: when dealing with text alone, as these applications do, rather than with a physically printed page, texts across numerous versions and editions can be quickly and easily compared.

The main differences between these applications is their presentation of information after collation. Figure 1 shows two very short witnesses entered into *CollateX*. In this case, there is only one variant—the word 'easily' has been added in the second text. In *CollateX*, where there is a variant, the path diverges. Using *Juxta*, information is displayed differently. Figure 2 shows the first paragraph of John B. Thompson's section on digital added value from *Merchants of Culture* in *Juxta Commons*, a web-accessible version of the application. Variants are highlighted, and clicking on the variants shows the nature of the variant, whether it is an addition, deletion, or substitution. Here, the example shows the change from 'seven' added values in the first edition to 'nine' in the second edition. Additional features offered by *Juxta* include side-by-side comparison and histograms.
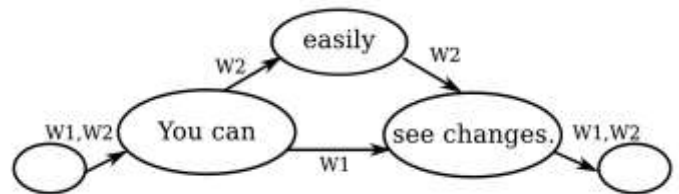

Figure 1: Two witnesses entered into *CollateX*.

Figure 2: A comparison of text from the first and second editions of *Merchants of Culture* in *Juxta Commons*.

---

[4] The origins of CollateX, a collaboration between the Beckett Digital Manuscript Project (BDMP) of the University of Antwerp and Huygens ING in The Hague, are described in Dirk Van Hulle's "The Stuff of Fiction."

The resultant studies making use of digital tools have the potential to further affect the field of textual criticism, which, according to G. Thomas Tanselle, is already experiencing a shift away from authorial intention, promoting instead "the forms of texts that emerged from the social processes leading to public distribution" (1). Even the term 'variant' has fallen under scrutiny, as it implies an original 'pure' text from which variation occurs, with the term 'rewritings' having been suggested as a possible replacement (Van Hulle 23).

In a blog post about her work on African American abolitionist Martin R. Delany's serial novel *Blake*, published between 1859–1861, Stephanie Kingsley describes her process of collation using *Juxta.* She writes:

> I uploaded the texts chapter by chapter into Juxta Commons, collated them, and then used Juxta's Edition Starter feature to produce HTML files which I then linked together into a navigable website created through GitHub. [...] The collations revealed many variants between the two serializations. Delany made scattered substantive changes to the work, usually changing one or two words or slightly [...] rearranging a sentence for clarity. The many small changes reflect an author interested in smoothing out and correcting his work but not in altering its original meaning. (Kingsley)

Kingsley's collation takes a traditional approach, but benefits from the use of digital tools, which enabled her to collate texts from two separate magazines and one reprint. However, born-digital texts with multiple authors can also be collated using these applications, as a 2013 examination of the continuous changes and updates on a single collaboratively-written article on Wikipedia explores (Schlosser). As these examples show, digital collation applications can aid researchers to both expand the boundaries of established practices as well as branch out into entirely new directions.

Visualizing Variants

Collation software creates visualization that, while astonishing in its speed and flexibility, is more suited to the researcher than the reader. The longer the text, and the more editions put under the metaphorical microscope, the more intricate the web of mapped changes becomes.

For his study of *Cloud Atlas*, Eve chose to visualize only the structural variations, or *syuzhet*,[5] using a Sankey flow diagram. He worked with pre-existing models, using the free software tool D3.js.[6] His resulting diagram (Figure 3) is read top to bottom, with the UK version

---

[5] *Syuzhet*, as Eve defines it, is "the way that a particular text organises its presentation of [the] narrative" as opposed to the *fabula*, "the chronological content of the narrative" (7).

[6] Eve has made his visualization software freely available. Moreover, he has invited other scholars to use his data toward a linguistic and/or textual comparison of the two versions of *Cloud Atlas*. See Eve 29.

represented on the left side and the US version on the right. Blocks represent narrative, with a thicker block indicating a greater length of structural similarity between the two versions. White space indicates content that is present in one version but not the other, the crossing of lines represents re-ordered content, and "where a block-link splits, this represents cases where one question was broken into several in the other edition" (Eve 9). [see Figure 3]

His focus on the structural differences rather than the textual alone differs greatly from the practices and concerns of textual criticism, as does his focus on version variants, the "changes in different printings of the same work" (Dedner 15). Here, it is not important what words exactly were changed but how and to what extent material was re-ordered, added, and subtracted. Eve's visualization abstracts the version variants for the chapter in question, giving a macro view that makes it possible to take in the entire chapter across two versions in a single glance.

While Eve presents a thoroughly academic study and his visualization functions accordingly, others have taken a more inviting, reader-focused approach. In 2009, designer and author Ben Fry mapped the first appearance of every word in all seven editions of Darwin's *On the Origin of Species*. Darwin's seminal work has a long publishing history, with the first edition published in 1859 and the

last published in 1872 (Fry). Each edition has substantial differences, made visible by Fry's map of the findings (Figure 4), titled *On the Origin of Species: The Preservation of Favoured Traces*. [see Figure 4]

The map is free to access online at the data platform *Fathom.* It is also interactive. Each edition is color-assigned, so it becomes immediately evident, for example, that a substantial amount of material was added by Darwin in the seventh and final edition, indicated by a large red column. By moving one's cursor over the map, color-coded text from the book appears, letting readers explore and read all the individual textual changes that were made. One can discover, for example, that the phrase "survival of the fittest" was not added until the fifth edition, and "evolution" first appeared in the sixth (Fry). The map, to quote Fry, "enabl[es] users to see changes at both the macro level, and word-by-word" (Fry).

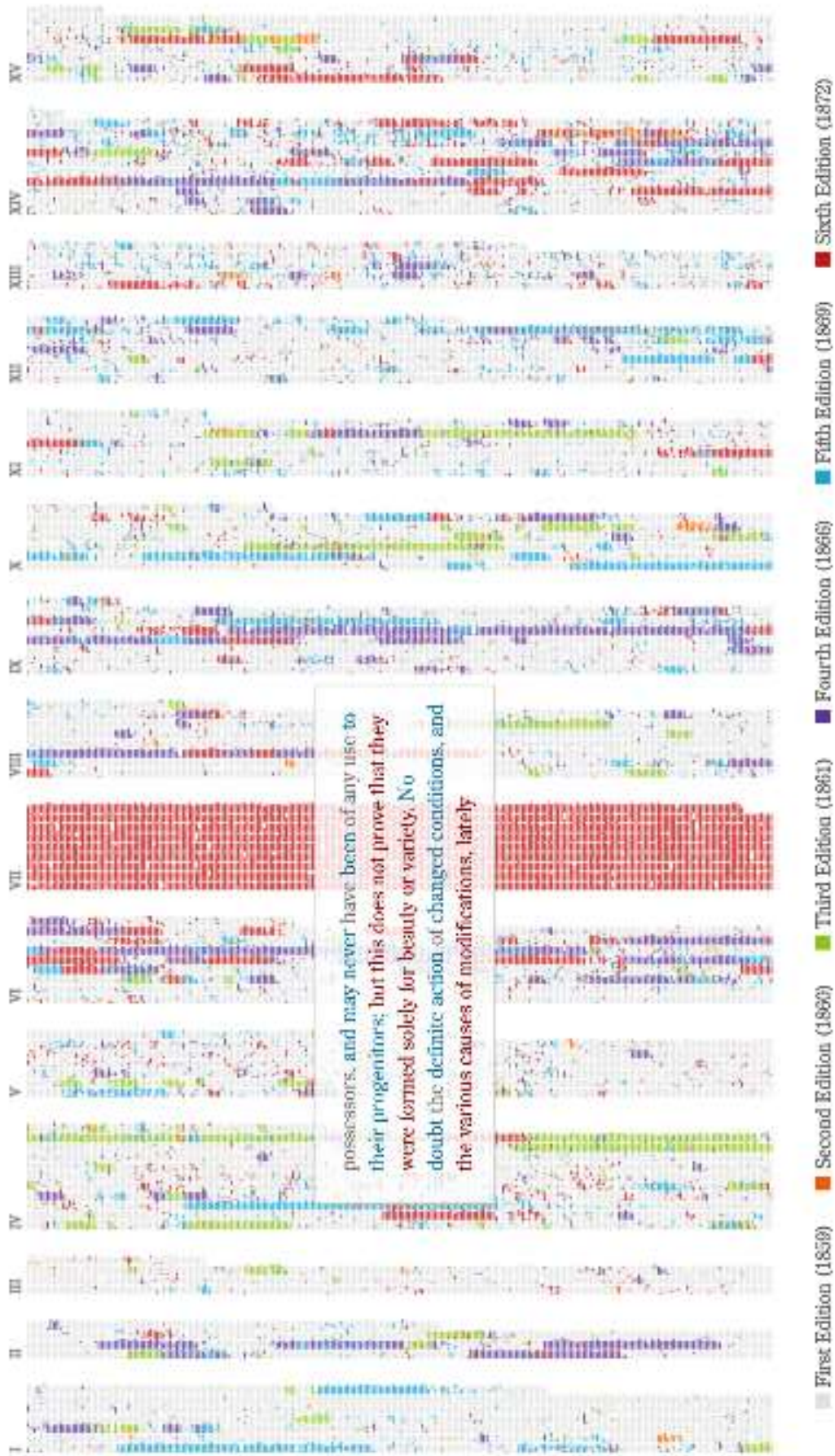Figure 3: Martin Paul Eve's visualization of the version variants in *Cloud Atlas.*

Figure 4: Ben Fry's interactive map of Darwin's *On the Origin of Species*.

Reading the Changes

Both Eve and Fry's visualizations make evident the storied writing and publication histories books have. What we might easily think of as "a book" is in fact a multitude of books that may or may not be identical in a variety of ways. These visualizations help to present the enormous variations our books, old or new, can have within their pages. What this all means for readers is the possibility to read, experience, and engage with texts in new ways: reading the changes that occur from copy to copy, edition to edition, or version to version.

Yet, as with the digital humanities more generally, digital collation is a budding field, and not all questions have answers. Is there a way, for example, to present collation results to readers in a way that is not inherently inclusive of all variants, but which rather guides readers through changes step-by-step, either chronologically or according to other parameters, such as stylistic changes versus error corrections? How could paratextual elements such as illustrations be represented? And how are scholars to deal with issues of copyright when working with and collating newer materials?

Even with such questions lingering, digital tools offer researchers and readers new methods for exploring the sometimes messy histories of books. Digital tools make it possible to present digitally collated material in invigorating new ways—interactive, reader-focused, and openly accessible—enriching reading experiences and allowing readers to trace the changes of some of their favorite books.

Bibliography

Biggs, Sarah J. "The Burden of Writing: Scribes in Medieval Manuscripts." *British Library Medieval Manuscripts Blog*, 03 Jun. 2014, https://blogs.bl.uk/digitisedmanuscripts/2014/06/the-burden-of-writing-scribes-in-medieval-manuscripts.html.

Dedner, Burghard. "Highlighting Variants in Literary Editions: Techniques and Goals." *Texts in Multiple Versions: Histories of Editions (Variants)*, edited by Luigi Giuliani, Herman Brinkman, Geert Lernout, and Marita Mathijsen, Rodopi, 2006, pp. 15–31.

Eve, Martin Paul. "'You Have to Keep Track of Your Changes': The Version Variants and Publishing History of David Mitchell's *Cloud Atlas*." *Open Library of Humanities*, vol. 2, no. 2, 2016, pp. 1–34.

Feather, John. *Communication Knowledge: Publishing in the 21ˢᵗ Century*. K.G. Saur Verlag, 2003.

Flood, Alison. "Cloud Atlas 'Astonishingly Different' in US and UK Editions, Study Finds." *The Guardian*, 10 Aug. 2016, https://www.theguardian.com/books/2016/aug/10/cloud-atlas-astonishingly-different-in-us-and-uk-editions-study-finds.

Fry, Ben. "On the Origin of Species: The Preservation of Favoured Traces." *Fathom*, https://fathom.info/traces/.

Hinman, Charlton. "Variant Readings in the First Folio of Shakespeare." *Shakespeare Quarterly*, vol. 4, no. 3, 1953, pp. 279–288.

Johns, Adrian. "The Book of Nature and the Nature of the Book." *The Book History Reader, 2ⁿᵈ Edition*, edited by David Finkelstein and Alistair McCleery, Routledge, 2006, pp. 255–272.

Kingsley, Stephanie. "grinding] gnashing: Juxta Collation of Delany's *Blake* Yields Insights into the Novel's Uncertain Conclusion." *Juxta*, 20 Jan. 2014, http://www.juxtasoftware.org/grinding-gnashing-juxta-collation-of-delanys-blake-yields-insights-into-the-novels-uncertain-conclusion/.

Mize, Britt. "Adam, and Chaucer's Words Unto Him." *The Chaucer Review*, vol. 35, no. 4, 2001, pp. 351–377.

Notaker, Henry. *A History of Cookbooks: From Kitchen to Page over Seven Centuries*. U of California Press, 2017.

Parsons, Scott. "The Text's the Thing: Using (Neglected) Issues of Textual Scholarship to Help Students Reimagine Shakespeare." *English Journal*, vol. 99, no. 1, 2009, pp. 85–90.

Schlosser, Emma. "Featured Set: Wikipedia Article on Benghazi Attack." *Juxta*, 9 Jan. 2013, http://www.juxtasoftware.org/featured-set-wikipedia-article-on-benghazi-attack/.

Tanselle, G. Thomas. "Textual Criticism at the Millennium." *Studies in Bibliography*, vol. 54, 2001, pp. 1–80.

Thompson, John B. *Merchants of Culture: The Publishing Business in the Twenty-First Century*. Polity Press, 2010.

---. *Merchants of Culture: The Publishing Business in the Twenty-First Century*. 2nd ed. Polity Press, 2012.

Van Hulle, Dirk. "The Stuff of Fiction: Digital Editing, Multiple Drafts, and the Extended Mind." *Textual Cultures*, vol. 8, no. 1, 2013, pp. 23–37.