Scholarship in interaction: case studies at the intersection of codework and textual scholarship

Zundert, J.J. van

Chapter 7

Apparatus vs. Graph: New Models and Interfaces for Text[1]

> "Interfaces are not simply objects or boundary points. They are autonomous zones of activity. Interfaces are not things, but rather processes that effect a result of whatever kind. For this reason I will be speaking not so much about particular interface objects (screens, keyboards), but interface effects. And in speaking about them I will not be satisfied just to say an interface is defined in such and such a way, but to show how it exists that way for specific social and historical reasons."
> (Galloway 2012)

7.1 Introduction

It is often said that textual scholarship is in transition (Kirschenbaum 2012), provoked by the increasing digitalization of society and culture. Yet perhaps textual scholarship itself is not so much in transition as is the basic "fabric" of its object of study, text. Almost all texts that are created today have at

[1] A previous version of this chapter has appeared as Van Zundert, Joris J., and Tara L. Andrews. 2016. "Apparatus vs. Graph: New Models and Interfaces for Text." In *Interface Critique*, edited by Florian Hadler and Joachim Haupt, 139:183–206. Kaleidogramme. Berlin: Kulturverlag Kadmos. The article is co-authored as a result of genuine collaborative thinking and research. Prof. Dr. Andrews provided most of the writing in the latter part of section 7.3 and most of the writing in section 7.5.

some point been produced or processed by a computer system, and hence almost all text today have also a digital existence, even if the final presentation is reliant on some physical medium. The field of scholarly editing has been struggling for over three decades to come to terms with how it could or should treat digital text, and to seek new contact points between the digital and the scholarly in the realm of text (cf. e.g. Landow 1994, Shillingsburg 2006; Robinson 2013a). These points of contact all rely on interfaces that allow textual scholars to interact with digital text. Any digitally informed transition to appear in textual scholarship should thus be tied to changing practices resulting from textual scholars interacting with texts by means of these interfaces. In this chapter a novel scholarly interface to text is presented that was developed in the course of research and development work by me and Tara L. Andrews. It will be argued that this interface presents text in a way distinctly different from the interfaces more commonly used in digital textual scholarship. The difference arises from applying a model for text that is at a remove from the well-known book or codex model. Arguably textual scholarship could enhance its engagement with text and its research capabilities by stepping back from the prevailing book model, and applying a more diverse range of models and interfaces.

7.2 The Computational Status Quo: Text-as-Sequence

The invention and development of the universal computer opened the way for the computational processing of textual information (Davis 2012). Although it is something of a foundation myth of digital humanities, the computationally aggregated concordance *Index Thomisticus*, a herculean effort by Father Roberto Busa first planned in 1946 with the cooperation of IBM and continuing into the 1970s, is usually seen as the first tangible result of textual computationality in the humanities (Winter 1999). Computational tractability of text has delivered powerful new means of curation and analysis to the hands of textual scholars. In textual analysis, stylometrics and stemmatology in particular were early "success stories" of computational approaches. Stylometrics is the – usually quantitative or statistically underpinned – analysis of literary style; its early applications included the identification of the au-

thor of a text based on its stylistic characteristics, and the field has expanded to investigate questions of genre, topic clustering, and narrative sequence. Stemmatology is the application of various methods, some based on statistics and others on deductive reasoning, to infer the order in which manuscript or print copies were derived from each other – somewhat akin to reconstructing a family tree of documents.

Our own earlier work provides examples of each. A stylometric measure known as Delta (Burrows 2002) was used in a case where two authors – one Penninc and one Vostaert, about neither of whom much is known – wrote successive portions of a medieval Dutch text, an Arthurian novel called *Walewein* (a Dutch analogue to the English "Gawain"). For reasons unknown Penninc did not finish work on the novel, and Vostaert later took it up, but there is no clear indication in the surviving text of where this transition occurred. Delta is effectively a measure for the use of high-frequency words. These are typically function words, also called functors in linguistics – generally small words with little lexical meaning, such as "is", "the", "me", and so on. The distribution pattern of this high frequency vocabulary is distinct for an individual author, and so acts as a sort of fingerprint. Applying an adaptation of this measure – which has become known as "rolling Delta" – to the *Walewein*, Van Dalen-Oskam and Van Zundert (2007) were able to get positive verification of the "handover" from Penninc to Vostaert (cf. figure 7.1). This type of research was only possible because of the application of computational statistics. Although the result could in theory have been derived by hand calculation, it would have been the work of many years to produce the millions of calculations needed.

The other example arises in the field of text stemmatology – the science of inferring a stemma (< Greek στέμματα, "pedigrees") of the copying relationship between manuscripts of an ancient or medieval text, based on evidence that can be found in the manuscripts available now. A "stemmatic method" bearing the name of the Classical and Germanic scholar Karl Lachmann was developed and refined over the course of the nineteenth century, and remains the standard for many textual scholars (Timpanaro 2005). Its use requires a collation of all available manuscript texts – a meticulous and formal detailed comparison of the text in its different forms – and depends on the

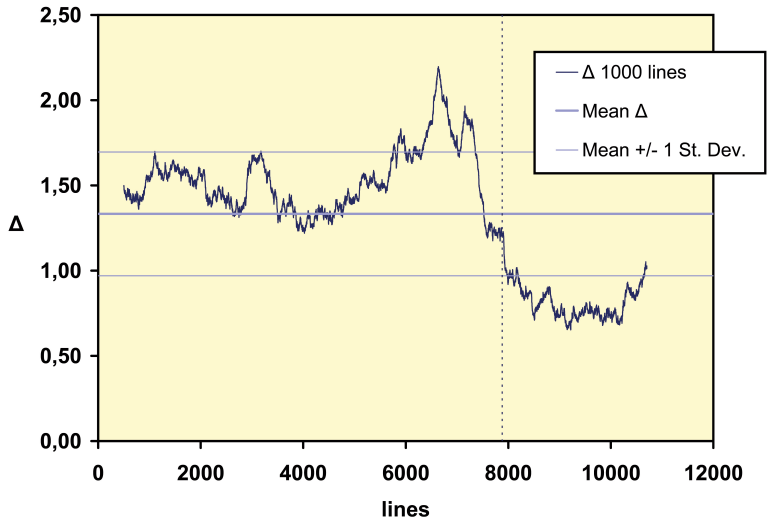


Figure 7.1: "Delta graph" showing changes in high frequency vocabulary through the text of Walewein, reproduced from Van Dalen-Oskam and Van Zundert 2007.

ability of the scholar to identify clear instances of copyist error that are unlikely to have been made twice.

Many textual scholars have argued that the Lachmannian method provides too much leeway for scholars to impart bias as to what is a "significant error" and how the stemma should thus be constructed. The method also runs into difficulty in the case of manuscript copies that were made by consulting more than one parent copy. Since the middle of the twentieth century the theory of Lachmann's method has been refined (Trovato 2014), but a competing methodology drawn from evolutionary biology has also been developed. This method, known as "cladistics", uses computational statistical methods to derive a genealogical tree according to the variants that are shared between subsets of manuscripts, with no need to distinguish scribal error in advance. Although cladistic methods arguably lend themselves more easily to computation, there have emerged in recent years a number of computational tools for creating stemmata semi-automatically according to Lachmannian principles. Examples include Roelli and Bachmann (2010), Camps and Cafiero (2014), and Barabucci, Di Iorio, and Vitali (2014). Many editors will employ a combination of methods to construct their stemmata; even so there remains some dispute in the field concerning which approach is more correct. In 2010 the Tree of Texts project began; its aim was to subject different methods of stemma creation to some form of empirical analysis. One result was a suite of tools for stemmatic analysis known as Stemmaweb.[2] The development of Stemmaweb required the creation of complex abstract computational models to represent text, its variation across manuscripts, and the "genetic" copying relation of those manuscripts to each other.

7.3 Function Follows Form: The Materiality of Text

Such computational analyses of texts and the statistical and digital tools used to perform them can be seen as new forms or modes of reading – "distant reading" as Franco Moretti (2013) calls it, as opposed to the customary "close reading". They can therefore be seen as new interfaces that allow readers

[2] https://stemmaweb.net/

and researchers to interact with text. Indeed Alexander Galloway defines interfaces not so much as objects or boundary points, but foremost as "processes that effect a result" (Galloway 2012:vii). In this sense these new digital or computational engagements with text are interfaces too: processes that allow a researcher to interact with text in a different mode, through a different process than by "reading the book". However, print text and the book, or codex, are to be regarded as interfaces by themselves as well. In everyday practice a text is identified with the book or another physical medium – journal, paper, etc. – carrying it, but in fact a book is just another way of engaging with the text. Even the text, understood as the semiotic system present on the pages, is just an interface. The actual text itself does not exist until and unless it is performed in some manner by an interpreter. Obviously the codex, especially in its function as a scholarly textual edition, has been an interface of paramount importance to the scholarly textual tradition.

Galloway argues that any interface that exists, exists for particular social and historical reasons. This is certainly the case with the scholarly print edition, where the use of the codex in textual scholarship and in scholarly print publications has been substantially shaped by the materiality of the page and the book. Although there exists a "language essentialist" line of reasoning, voiced by Thomas Tanselle and still widely adhered to, which holds that a text can be equated to its words, we also find scholars such as Jerome McGann who argue that the materiality of literature is an essential aspect of its content (Welch 2010). Looking at William Blake's illuminated lyrical work (figure 7.2) for instance, it becomes difficult to reject the influence that materiality has over the performance, experience, and interpretation of a poem's text. Blake's poem is a visual experience that goes beyond the characters and words. The text itself is not typeset, but engraved; this allows the textual and visual elements to flow together and emphasize each other in ways impossible to achieve through conventional typesetting. Through the poem's material properties Blake brings textual and visual elements into a "dialogue" that provides a different aesthetic experience, and therefore perhaps also a different or enhanced interpretation of the work – an experience that cannot be reduced to standard movable type.

However, that is often exactly what happens when works are remediated in the digital realm. Figure 7.3 depicts the remediation of the same poem

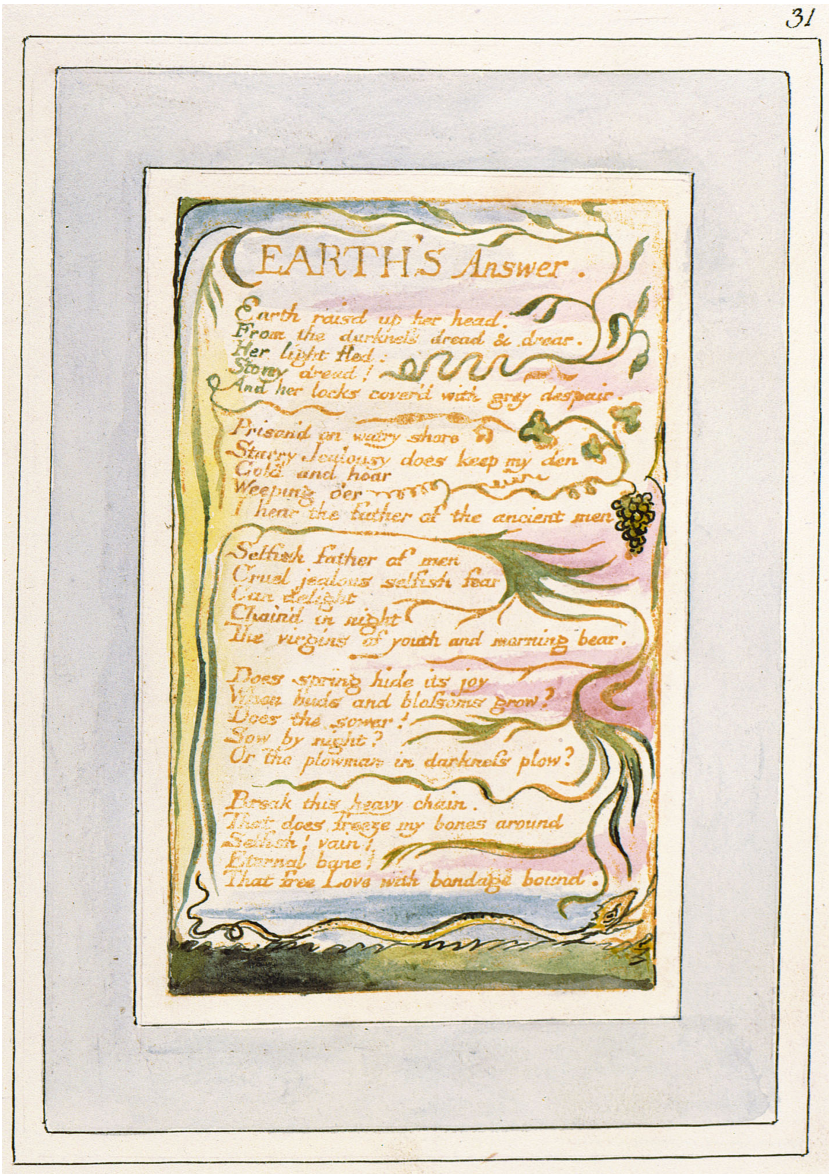


Figure 7.2: Songs of Innocence and of Experience, copy V, 1821 (Morgan Library and Museum) object 31 Earth's Answer (Public domain, via Wikimedia Commons).

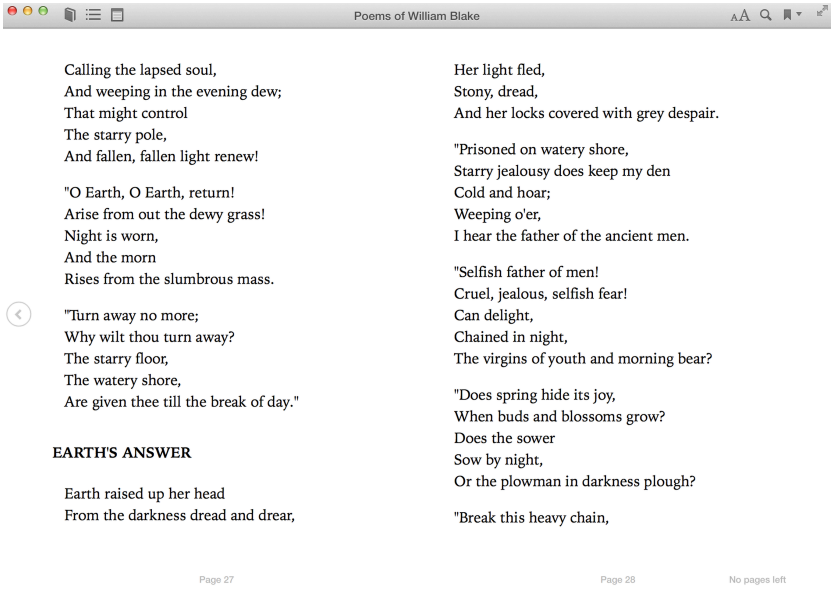


Figure 7.3: Poems of William Blake, p.27 of the Project Gutenberg eBook.

through one of the ePubs that are available via Project Gutenberg (Blake 1996[1821]). This version of the poem has what could be called the "Tanselean" essence of the text, which provides for a totally different experience, a distinctly different aesthetic. What makes this re-representation of Blake's text so different from its original is amongst others that it is offered as a machine readable text, which means that each glyph in the text is encoded as a number. Each number maps by human agreement – in this case the Unicode standard – to a character in an alphabet that is meaningful to human interpreters. It is this encoding which makes any letter "a" formally comparable to any other "a" in the computational or digital realm. This formalization of characters is what makes text computationally tractable, but it also causes the machine readable character to be reductive. The distinct material features that make Blake's poem unique are reduced to computationally comparable features. Text in the computational realm is likewise generally reduced to a "string": a linear series of characters and words that enabled – or rather were necessitated by – Morse code, telegraphy, Shannon's information theory and the Turing model (Shannon 1948; Davis 2012). This linear encoding was arguably reinforced and amplified by the severely limited memory and storage capabilities in the early days of computing. This formalism of linear information representation is both very powerful and very lossy: powerful because it allows for such stylometric and stemmatic analyses as we have shown above, but lossy with respect to the distinctive materiality that is characteristic of many real-world print texts. Yet as we have seen, the materiality of text is essential to textual scholarship and especially to scholarly editing, as it is directly connected to modes and possibilities of interpretation. Editions can only ignore the dialogue between the text under consideration and the materiality of its medium at the peril of reducing the "interpretational space" the original rendering of a text provides for.

7.4 Evolution of Scholarship in the Margins

There is a strong self-reflective relation between the materiality of the book and textual scholarship. The very materiality of the manuscript codex, and later the print publication, have shaped the formalisms of textual scholar-

ship, and thus also in part the ways texts and documents are studied within textual scholarship. This can be gauged from a ninth century manuscript as depicted in figure 7.4. This page shows the early traces of how scholarship formalized itself – quite literally – in the margins of manuscript documents (Teeuwen 2011). The main text, supposedly a faithful rendering of an original text of Martianus Capella, forms the main "body" of the text on the page. Ninth-century scholars used the margin of the codex to squeeze in as many additional annotations, additional knowledge, and clarifications, as the space on the vellum would allow. Glosses were written between the lines, and Tironian notes (an ancient and medieval shorthand system for professional scribes) were used to compress even more information into the document. All these notes, annotations, glosses, and clarifications allowed the scholars to network their knowledge, to allow them to testify to the interaction between their thinking and the writing of their predecessor Martianus Capella (and, for that matter, the thinking of the various commentators before them).

The materiality of the medium – vellum and manuscript – has been pivotal in shaping how this dialectic and discourse took to paper. The margin was not invented especially for note taking, but it was there and came to be used for that purpose. Eventually the tradition of adding notes in the margins became more and more formalized and by the nineteenth century had evolved into what we now call "the apparatus" – a complex piece of scholarly technology to express as much scientific information about the text as possible in a confined space. Figure 7.5 shows a page from a typical modern critical edition of a classical or medieval text. The purpose of a critical edition, in most cases, is to present what, in the editor's judgment, is the most correct version of the text that can be deduced. In order to do this, the editor must collect as many manuscript copies of the text as can be found, compare them word by word to discover the differences between them, perform a stemmatic analysis to understand the approximate history of the text transmission, and finally construct a text based on all this evidence that is most closely in line with the editorial principles that the editor uses. In the critical edition, that text is presented at the top of the page, usually with line numbers for reference. The remainder of the page is taken up with a set of specially formatted footnotes (the apparatus) that encode, ideally, all of the information that the editor

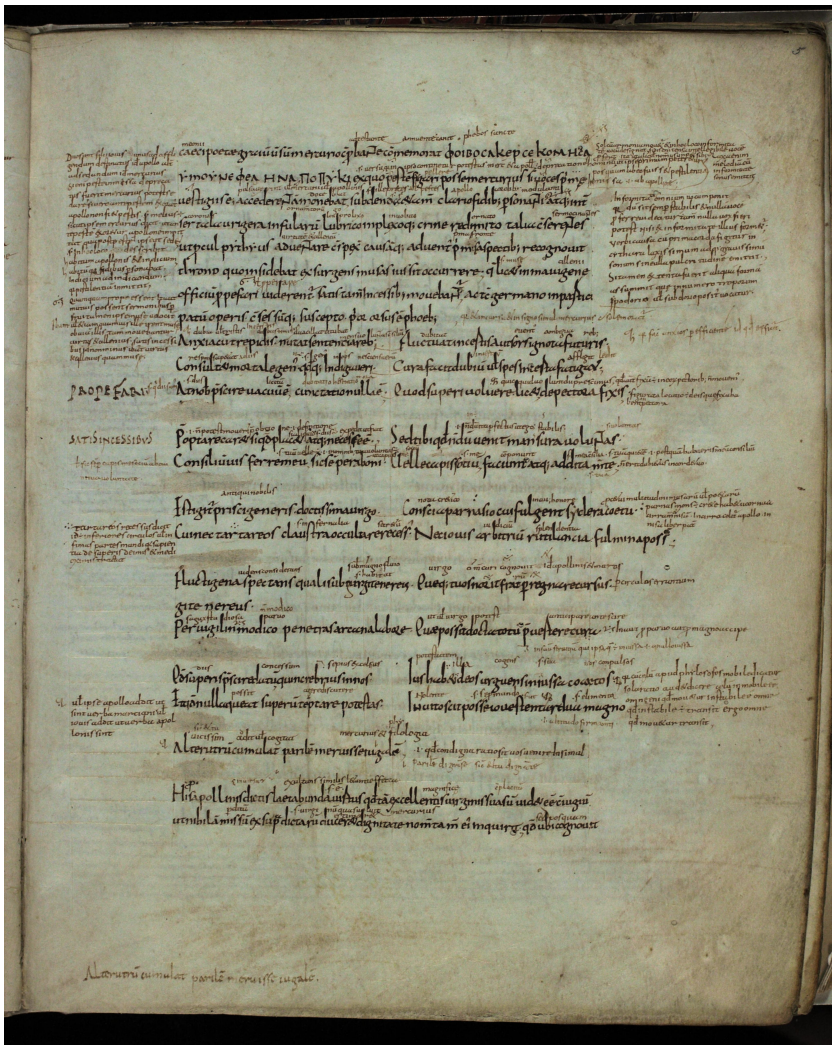
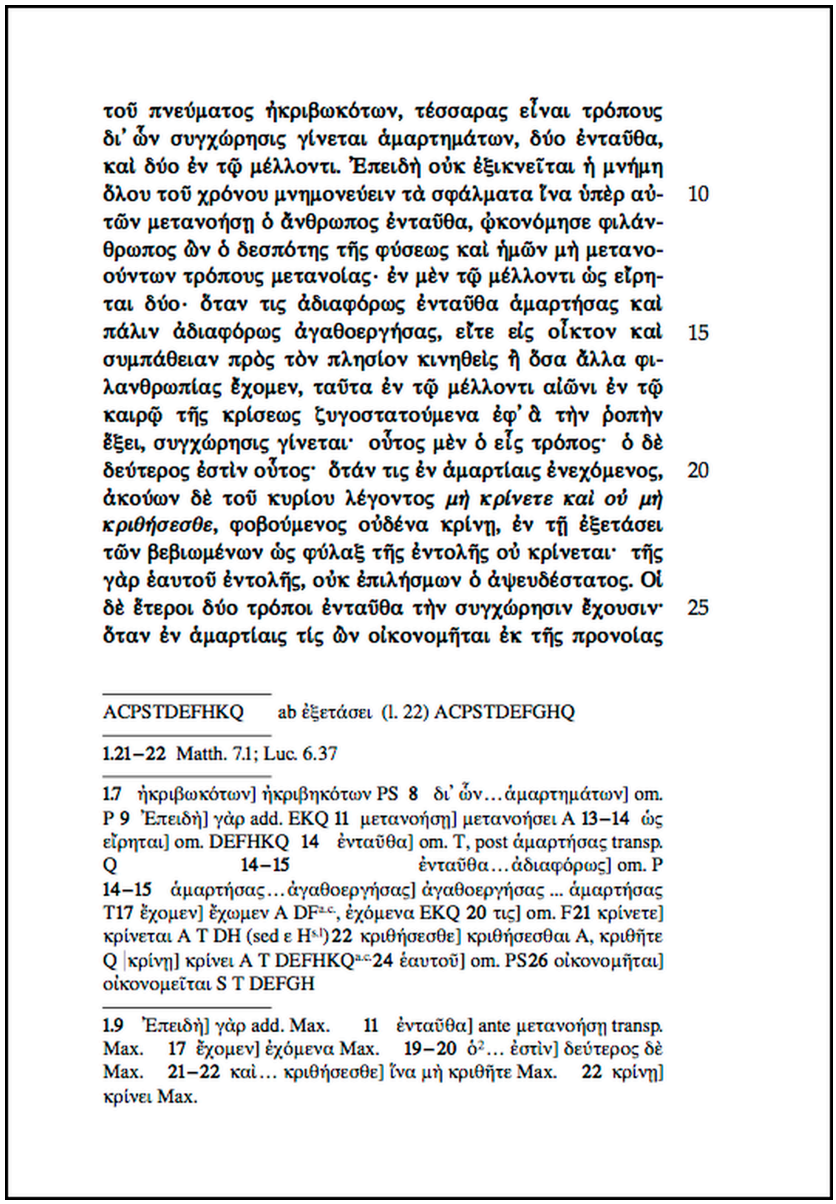


Figure 7.4: Leiden, University Library, Voss.Lat.F. 48, fol.5r.


τοῦ πνεύματος ἠκριβωκότων, τέσσαρας εἶναι τρόπους
δι' ὧν συγχώρησις γίνεται ἁμαρτημάτων, δύο ἐνταῦθα,
καὶ δύο ἐν τῷ μέλλοντι. Ἐπειδὴ οὐκ ἐξικνεῖται ἡ μνήμη
ὅλου τοῦ χρόνου μνημονεύειν τὰ σφάλματα ἵνα ὑπὲρ αὐ- 10
τῶν μετανοήσῃ ὁ ἄνθρωπος ἐνταῦθα, φκονόμησε φιλάν-
θρωπος ὢν ὁ δεσπότης τῆς φύσεως καὶ ἡμῶν μὴ μετανο-
ούντων τρόπους μετανοίας· ἐν μὲν τῷ μέλλοντι ὡς εἴρη-
ται δύο· ὅταν τις ἀδιαφόρως ἐνταῦθα ἁμαρτήσας καὶ
πάλιν ἀδιαφόρως ἀγαθοεργήσας, εἴτε εἰς οἶκτον καὶ 15
συμπάθειαν πρὸς τὸν πλησίον κινηθεὶς ἢ ὅσα ἄλλα φι-
λανθρωπίας ἔχομεν, ταῦτα ἐν τῷ μέλλοντι αἰῶνι ἐν τῷ
καιρῷ τῆς κρίσεως ζυγοστατούμενα ἐφ' ἃ τὴν ῥοπὴν
ἕξει, συγχώρησις γίνεται· οὗτος μὲν ὁ εἷς τρόπος· ὁ δὲ
δεύτερος ἐστὶν οὗτος· ὅτάν τις ἐν ἁμαρτίαις ἐνεχόμενος, 20
ἀκούων δὲ τοῦ κυρίου λέγοντος *μὴ κρίνετε καὶ οὐ μὴ*
κριθήσεσθε, φοβούμενος οὐδένα κρίνῃ, ἐν τῇ ἐξετάσει
τῶν βεβιωμένων ὡς φύλαξ τῆς ἐντολῆς οὐ κρίνεται· τῆς
γὰρ ἑαυτοῦ ἐντολῆς, οὐκ ἐπιλήσμων ὁ ἀψευδέστατος. Οἱ
δὲ ἕτεροι δύο τρόποι ἐνταῦθα τὴν συγχώρησιν ἔχουσιν· 25
ὅταν ἐν ἁμαρτίαις τίς ὢν οἰκονομῆται ἐκ τῆς προνοίας

ACPSTDEFHKQ ab ἐξετάσει (l. 22) ACPSTDEFGHQ

1.21–22 Matth. 7.1; Luc. 6.37

1.7 ἠκριβωκότων] ἠκριβηκότων PS 8 δι' ὧν…ἁμαρτημάτων] om. P 9 Ἐπειδὴ] γὰρ add. EKQ 11 μετανοήσῃ] μετανοήσει A 13–14 ὡς εἴρηται] om. DEFHKQ 14 ἐνταῦθα] om. T, post ἁμαρτήσας transp. Q 14–15 ἐνταῦθα…ἀδιαφόρως] om. P 14–15 ἁμαρτήσας…ἀγαθοεργήσας] ἀγαθοεργήσας … ἁμαρτήσας T17 ἔχομεν] ἔχωμεν A DF$^{a.c.}$, ἐχόμενα EKQ 20 τις] om. F21 κρίνετε] κρίνεται A T DH (sed ε H$^{s.l.}$) 22 κριθήσεσθε] κριθήσεσθαι A, κριθῆτε Q |κρίνῃ] κρίνει A T DEFHKQ$^{a.c.}$ 24 ἑαυτοῦ] om. PS 26 οἰκονομῆται] οἰκονομεῖται S T DEFGH

1.9 Ἐπειδὴ] γὰρ add. Max. 11 ἐνταῦθα] ante μετανοήσῃ transp. Max. 17 ἔχομεν] ἐχόμενα Max. 19–20 ὁ^{2}… ἐστὶν] δεύτερος δὲ Max. 21–22 καὶ… κριθήσεσθε] ἵνα μὴ κριθῆτε Max. 22 κρίνῃ] κρίνει Max.

Figure 7.5: Extract of critical edition of Florilegium Coislinianum β′. This particular extract was taken from De Vos et al. (2010), it was rendered for illustration by T. Andrews using the Classical Text Editor software.

used to arrive at his or her conclusions.

This critical edition uses four separate apparatuses. In them, each manuscript that was used for the edition is represented by a sigil – an identifier such as A, B, C, etc. The first block holds the apparatus siglorum, which is a list of the manuscripts that were consulted for this particular span of text. In this case the reader can see from the apparatus that manuscript G was included beginning from the word ἐξετάσει on line 22. The second block notes references to the Bible; the fourth notes parallels with a different but related text. It is the third block, the apparatus criticus, where the majority of the editor's work is put on display. Each entry in this block gives a line number, a word that appears on that line of the text, and a list of the alternatives that appear in one or more manuscripts. Thus, for example, the first entry in the block

```
1. 7 ἠκριβωκότων] ἠκριβηκότων P S
```

tells the reader that, where ἠκριβωκότων appears on line 7 of this page, the manuscripts known as P and S read ἠκριβηκότων at that point instead. The next entry

```
8 δι' ὧν...ἁμαρτημάτων] om. P
```

tells the reader that the text on line 8 from δι' ὧν to ἁμαρτημάτων inclusive does not appear in (is omitted from, thus "om.") manuscript P. Abbreviations such as "om.", usually in Latin, appear fairly commonly in a critical apparatus, and the apparatus entries themselves can be arbitrarily complex in order to conserve space, which is at a premium in most printed texts. For example, beginning on the sixth line of the apparatus the reader encounters the entry

```
21 κρίνετε] κρίνεται A T DH (sed ε Hs.l.)
```

Here the reader is to understand that the word κρίνετε in line 21 appears in manuscripts A, T, D, and H as κρίνεται; that D and H are considered to be members of a related group of manuscripts (thus the lack of space between the sigla); and that H also contains the κρίνετε form written in above the line (s.l. = "supra lineam"). It is left to the reader to make the most probable interpretation of this fact – that the scribe of H initially wrote κρίνεται, and then corrected the text to read κρίνετε – although without either a more explicit indication or an image of the manuscript itself, the reader cannot be sure that this is what the editor meant to convey.

7.5 The Fear of Digital Freedom

The pair of examples above serves to show how the materiality of the codex – the technology of the book – shaped the practice and the technical apparatus that textual scholarship to this day applies in its publications, or in other words: how it shaped the interfaces textual scholarship creates to texts. There was in a certain sense a "dialectic" between materiality and intended purpose that mutually shaped the practice of scholarly commentary. Digital technology radically changes the nature of that dialectic. This is not, however, because digitality does away with materiality. At the most basic level, after all, bits of digital information are represented as material items: either as an analogue (electrical) signal within the most basic element of a computer's technical structure (the flip-flop, which is the simplest electronic switch architecture that can represent a dichotomy by keeping an electric current in either one of its two tiny circuits but not both), or in some more perpetuating material form (e.g. as tiny magnetized regions on a hard drive). As such there is little that is non-material about digital information (cf. for instance Kirschenbaum 2008). But what digital computation produced, propelled by the mathematical and technological developments by inter alia Turing and Shannon, were the most atomic elements of information, literally the bits, commonly understood as ones and the zeros. These most elementary atoms of computation allow us to derive by abstraction and composition any formal logic and language we can devise. The computer thus delivers an all but infinite freedom to create logic spaces and visual representations with and of

these. One might regard computer code as an ultimate form of Lego: you can design your own blocks and even your own rules on how to stack them, unlimited and unbounded.

For interfaces and the creation of visualizations, this allows a radical freedom of expression. Suddenly anything that is imaginable as a visualization, as an interface for text within textual scholarship, may be realized. The limits of the material page, of the book, need no longer be the limits that guide the thinking about visualizing and interfacing with texts in textual scholarship. To a field such as textual scholarship, which – given one of its functions as the agent of preservation of texts – is consciously somewhat conservative in its methodology, this can actually be a disturbing development, especially in its function of philology. As Jerome McGann points out, philology is concerned with the memory of humanities, the record of humanistic expression – in what could be called almost a desperate attempt to stabilize that record (McGann 2013). Digitality in its radical freedom of expression and unlimited versatility of visual form, is obviously not a medium that allows for a convergence of method guided by materiality. This is not because of a perceived (but imaginary) loss of materiality, but rather because of the removal of the material limits of the page that influenced such a convergence of practice and expression. To a field that is concerned mostly with stabilizing its objects of study this may be more than a little disconcerting. In what could be a backlash in reaction to this radical freedom, some textual scholars seem to argue for a return and a reconfirmation within the digital environment of the structures that govern the book text, arguing that digitality should primarily provide a straightforward remediation of the book/text-structure in a digital environment, thus offering the digital text as a faithful simulacrum of the material and textual structures present in a material exemplar (Shillingsburg 2006; Pierazzo 2011). In which case the material limits of a physical medium are consciously superimposed onto a digital medium to limit its radical freedom.

Paradoxically then, digitality, which in principle suggests the liberation of material limits, currently serves primarily to reaffirm these limits. Moreover it seems even to narrow these limits and perhaps to reduce the expressiveness of textual materiality, as in the case of the visual poetry of William Blake and the inability of machine-readable text to capture and express its materiality.

However, rather than experiencing the radical freedom of the digital environment as a threat to the stabilizing functions of philology, we suggest that this freedom should be seen as an opportunity to explore the alignment of visualizations and interfaces of texts with the specific functions of a text. To support this we must return to an axiom in textual scholarship which holds that no text is actually stable. Rather, text is fluid. It is volatile: it gets changed, adapted, it crosses medium boundaries, it interacts with other texts, and so on (cf. for instance Bryant 2002; Coffee et al. 2012). Similarly an edition of a text, be it a scholarly one or not, is not to be equated with the actual text itself. An edition is always a re-illumination, a re-edition, a re-issue of a text. Due simply to the fact that a textual scholar had to interpret the text in order to produce it, the text is not the same as the text the scholar found in his or her witness(es). Because it has been interpreted, and has been given a new form by the textual scholar, it is by definition a new text. In the words of Bernard Cerquiglini (1999): an edition is not the text itself, an edition is an argument about that text. If a scholarly edition – or for that matter any edition – indeed is an argument about a text rather than the text itself, then it follows that the form should fit the argument. Of course this can mean that the edition should mimic as best as it can the "original" text, insofar as the scholarly editor chose to focus on the preservation aspect or conservational purpose of a digital re-rendering of the text. But utilizing the radical freedom of expression of the digital environment to the benefits and purposes of textual scholarship could also mean trying to fit a specialized interface to each of any perceived functions of the argument. Thus for instance, if the scholar is arguing the genesis of the text, the reader or user of the edition should be able to gauge the edition as if the text was being written: why indeed not animating or simulating the argued writing process, if the writing process is the aspect a scholar is scrutinizing? Then again, if the editor is arguing the reading of the text, we might simply want a clean reading text, bereft of any annotation or apparatus; perhaps a simple ePub text that can be read on a tablet. In other words an editor should be able and expected to create an interface fit for the purpose of the argument being set forth about the text.

Interfaces to text can in this sense offer different views on text, highlighting specific aspects that in traditional interfaces may be less pronounced or even impossible to express clearly. In computer science terms this is called many vi-

sualizations for one model (cf. for instance Gamma et al. 1994:14). In textual scholarship – admittedly abstracting away for a moment a number of text-philosophical subtleties – it can be called many instantiations or functions of a work. (For the relation between work and instantiation cf. Richard P. Smiraglia (2001:66); on the functions of text cf. Sahle (2013).) Deriving multiple views from one model will require something far more advanced than the string model however. The linear data structure of the string requires that too many additional features of a text be bound to it by other digital objects. Usually this results either in document markup schemes – such as XML[3] or Text Encoding Initiative's TEI-XML[4] – or an object model defined in programming code. Neither of these solutions is satisfying. Though an object model allows for maximal flexibility in expressing the structure, semantics, and commentary of a text, it lacks the attractive features of easy exchange and a data structure that concisely integrates all knowledge about a text, like a markup language such as XML does. Markup languages on the other hand lack the flexibility of compositional models and are unsuited to deal with multiple hierarchies, which means they are infeasible as a solution for expressing different views or differing interpretations of a text and its structures, as these are typically tied to or resulting in different markup hierarchies (cf. Sperberg-McQueen 2002; Schmidt 2014).

Between markup at one end of the solution space and object models at the other, graph structures seem ideally positioned as a middle ground. They do allow for expressing the complicated composite relations between objects that lend object models their power and versatility of modeling – and thus they are flexible and expressive enough to capture the multidimensional, multi-layered, composite relations between semiotics, structure, semantics, annotations, commentary, and various other layers of text. On the other hand they are integrated, self contained data structures much like XML hierarchies, easy to exchange and port for interoperability. In short we argue that graphs are ideally positioned to attenuate the reductive effect of basic linear computational strings, while avoiding some of the side effects of hierarchical markup structure.

[3] Exstensible Markup Language, see http://www.w3.org/XML/.
[4] http://tei-c.org/

Given these properties it is hardly surprising that graphs – as a data model – have found applications meanwhile in textual scholarship, most notably in the realm of describing and expressing textual variation (Schmidt and Colomb 2009; Haentjens Dekker et al. 2014). Indeed we use graphs as a model for text variation in our own work. However, we wondered if the graph itself could not serve as a far more intuitive and helpful "map" for variants description than the difficult to produce and difficult to use apparatus that we showed above. Thus we devised a visualization that depicts the graph itself to chart textual variation between different witnesses of the same text or work. We have also found this variant-specific visualization and interface to be a useful interactive tool for textual scholars.

7.6 Text and Its Variation as a Graph

The guiding principle of a variant graph is that all versions of the text have a beginning, and an end, and each version has a particular sequence of words that corresponds with the sequence of other versions. The simplest variant graph, of a single text, might look like the one depicted in figure 7.6 representing a popular sentence used for testing a typewriter.

Figure 7.6: Graph for a single text.

This fairly common sentence does have a common variation in popular culture that one might hear spoken. If we were to express that variation according to the rules of the printed apparatus criticus, we might say

```
dogs] sleeping dog *Oral trad*.
```

We can express that substitution in a graph as depicted in figure 7.7.

Figure 7.7: Substitution expressed in a graph.

Here our graph edges (the arrows) are labelled with the respective sources of our sentences – in this example, one comes from a written text ("Text") and the other from an oral tradition ("Oral"). Where the versions diverge, so too do the labels. By following the labels for any named version through the graph, the reader/user can reconstruct that version, unambiguously, in its entirety.

In the terminology of textual scholarship, we refer to each of the green nodes in the graph as a "reading". The labels on the edges refer to individual versions of the text, known as "witnesses". Readings that occur in corresponding ways in different witnesses, such as "dogs" versus "sleeping dog" in our example, are known as "variants", and it is these variants that the graph was developed to display. The variant graph is thus a means of expressing the relationships between several copies of the same text in a unified way, giving the reader an interface to the text that provides several immediate advantages over the apparatus criticus. The first is that the reader can quickly grasp the magnitude of the variance and fluidity of a text in a visually intuitive way. The second is that this interface refrains from inherently privileging any single version of the text; rather, the reader is free to consider the various textual copies as a collective entity, independent of any structure imposed by the author of the apparatus criticus.

The third advantage of the variant graph is that it can also be used to express more explicit information about the textual variation than is typically included in any printed edition – for example, the specific relationship that might exist between a pair of variants. To illustrate this, we can turn to the Greek text used in our apparatus criticus examples above. Our apparatus entry from line 7

```
7 ἠκριβωκότων] ἠκριβηκότων P S
```

becomes a simple division as depicted in the graph of figure 7.8.

Figure 7.8: Simple division expressed in a graph.

with a red line to join the variant readings that, in this case, indicates that these are spelling variations of the same word. In this example we also see that the thickness of the lines in the text path is used to give an impression of the respective number of witnesses that follow each sequence.

The omission of the phrase recorded in our apparatus criticus for line 8 of the text

```
8 δι’ ὧν...ἁμαρτημάτων] om. P
```

can be represented in two ways. An initial graph might look as shown in figure 7.9.

Figure 7.9: Graph expressing an omission in witness P.

A textual scholar would most likely object to such a representation – she or he would argue that this graph shows the absence of five separate readings in P, suggesting five separate acts of omission by its scribe. A better graphical representation of the text might thus be constructed as in figure 7.10.

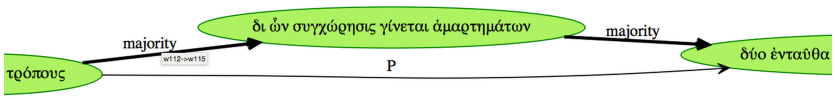


Figure 7.10: Graph expressing omission in witness P as singel reading.

In this case the variant is understood to be a single reading, comprising the entire phrase, that was omitted in a single act.

A more complex example from the original apparatus criticus is the following, which required three entries to describe in full.

```
14 ἐνταῦθα] om. T, post ἁμαρτήσας transp. Q
14–15 ἐνταῦθα...ἀδιαφόρως] om. P
14–15 ἁμαρτήσας...ἀγαθοεργήσας] ἀγαθοεργήσας ... ἁμαρτήσας T
```

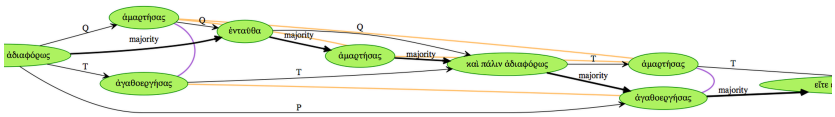


Figure 7.11: more complex variation in a graph.

Expressed as a graph (figure 7.11) the purple lines represent readings that have parallel grammatical structures, and the orange lines represent readings that are the same word, transposed to another location in the text. The three distinct apparatus criticus entries are thus unified into a single representation of the overall textual fluidity.

Although the graph interface to the text has the great advantage that no single version of the text need be prioritized, it is also possible to use color as an additional channel of information. As the editor studies a text, she or he may well wish to construct a "red thread" through the possible readings,

marking out the argument to be made concerning the sequence of readings that might constitute the canonical text. This can be marked, quite literally, in red (figure 7.12).

Figure 7.12: A 'red line' of edges marks a canonical reading.

7.7 Conclusion

Textual scholarship is rooted in a tradition where many scholarly functions demanded of a text were made part of that text within the material confines of the codex. Those material limits and physical context of the codex essentially converged with the conceptual demands of scholars to shape a highly technical and complex apparatus that combined the many functions that scholars wished to see represented in a text. The transmediation of text into the digital medium has so far been problematic from the perspective of textual scholarship, as the typical digital and computational forms have not provided a very welcoming environment for text. On the one hand, the basic and far too simplistic structure of the string severely denies the complexity of text. On the other hand, the radical freedom of computational modeling presents textual scholarship with a vastness of unexplored presentational possibilities that can be more terrifying than reassuring to a discipline that often concerns itself with the conservation and stability of textual information. At the same time, modern computational methods of analysis clearly demonstrate a need and a use for new forms of representation and interface with text in the digital environment. In this chapter it has been suggested that a recently-developed basic representation of text – the text-as-graph model – strikes a useful middle ground between the computational and the physical representations of text, providing as it does for both conceptual complexity and computational tractability. We demonstrated its use by application to

one of the most common needs within textual scholarship: a representation of the variation of a text. Whereas textual variation is usually expressed in print by a dense and intricate apparatus criticus that appears together with other forms of apparatus in accompaniment to a presentation of the canonical text, through the text-as-graph model the text can instead be made tangible through an innovative graphical interface focused specifically on this one aspect. From this work we draw the conclusion that, rather in contrast with the converging practice of digital textual scholarship as a remediation of the codex, the potential of digital interfaces should be explored by developing an array of specific interfaces for specific textual functions, in accordance with views on (digital) textual scholarship as a form of disciplined play (Rockwell 2003). In this way, textual scholarship can embrace the richness and disruptive freedom of the digital medium as a positive force for knowledge, rather than an externally-imposed threat.