The Visual Concordance: The Design of Eye-ConTact¹

Dr. Geoffrey Rockwell McMaster University grockwel@mcmaster.ca

Abstract

Computer-assisted text analysis tools are used, but rarely theorized which makes it difficult to asses the hermeneutical role of textual visualization. This paper proposes that text-analysis tools represent a text or corpus by rearranging the text according to a question by the user. This proposal is used to discuss the place of textual visualization systems as an alternative way for the user to ask questions of a text through software. This paper is not so much a reflection on the results of the visualization of text, as a reflection on the purpose of text-analysis tools in order to understand visualization. To do this the paper starts with a provocation from Lucian of Samostata regarding monstrous texts and reminding us that the problem of hybrid texts is not new. From there the paper summarizes the history of text-analysis tools leading to comments on the hermeneutics of concordances in order to finally discuss the design of a prototype visual text-analysis tool called Eye-ConTact. We can only understand the place of visualization environments in textual studies if we examine the nature of the concordance and the tools that evolved from it.

Provocation

Lucian of Samostata, the writer of dialogues of the dead, sets up a series of trials in his dialogue "The Double Indictment" two of which deal with Lucian's style of writing. In one of these trials the character Dialogue takes a Syrian (Lucian himself) to task for corrupting him.

"What is most monstrous of all, I have been turned into a surprising blend, for I am neither afoot nor ahorseback, neither prose nor verse, but seem to my hearers a strange phenomenon made up of different elements, like a Centaur." (Lucian, "The Double Indictment", p. 147)

Dialogue's complaint is that Lucian has adapted him from philosophical use to satirical use, creating a monster or hybrid genre that is neither horse nor man. Needless to say in Lucian's dialogue the Syrian is found innocent of corruption, unlike Socrates.

In this paper I want to talk about how text-analysis tools corrupt texts and produce hybrid or monstrous new texts. This paper is not so much a reflection on the visualization of

¹ I would like to thank the McMaster University Arts Research Board and the McMaster Computing and Information Services for their support. Patricia Monger from CIS did the programming in Visual Basic of Eye-ConTact. Many of the ideas presented here grew out of conversations with John Bradley.

text, as a reflection on the purpose of text-analysis tools in order to understand visualization. To do this I started with a provocation from Lucian of Samostata. From there the paper will briefly summarize the history of text-analysis tools leading into comments on the hermeneutics of concordances in order to finally discuss the design of a prototype text-analysis tool called Eye-ConTact. The point here is that we can only understand the place of visualization environments if we examine the nature of the concordance and the tools that evolved from it.

History of Text Analysis Tools

The first text-analysis tools were designed to create paper concordances. Father Roberto Busa in the late 1940s was one of the first to use of computers in the production of concordances with his *Index Thomisticus*, a project that began by using index cards, moved onto analogue information technology in the 50s and migrated to the computer. The results were finally delivered in the 1970s with a CD released in 1992.²

The concordance, however, goes back to the 13th century. Hugh of St. Cher is credited with directing the production of a concordance to the Vulgate bible by brother Dominican monks in Paris. This concordance, supposedly finished by 1247, suffered in that it only had references not quotations that give a sense of context. Quotations were apparently added later by English Dominicans to a concordance that has not survived. Finally, a concordance attributed to Conrad of Halberstadt improved on the model, leaving us by the end of the century with a concordance that provided some context along with references. It is worth noting that the Latin word "concordantiae" is in the plural; our concordance is really a collection of concordances, one for each word. Each individual concordance is an aggregation of passages under a key word that agree or concord in some way.

To return to my brief history of text-analysis tools, the first generation of available tools were batch tools that were not interactive, but were designed to produce paper concordances. This can be seen in the names and operations of many of these tools. COCOA stands for Count and Concordance generation on the Atlas, though some of suspect it was the drink of choice of the programmers.³ The Oxford University Computing Service took over COCOA in 1978 and produced OCP and later Micro-OCP. The OCP stands for the Oxford Concordance Program. TACT itself stands for Text-Analysis and Concordance Tools.

With the availability and increasing power of micro-computers in the 1980s Text-Analysis tools migrated from mainframes down to personal computers. OCP led to Micro-OCP and new programs came out for the personal computer like the Brigham

_

Concordances.

² As Joseph Raben recounts in "Humanities Computing 25 Years Later", "The importance of the computer as a lexicographical tools was apparently first demonstrated in a computer-assisted study of Thomas Aquinas conducted by Roberto Busa, a Jesuit teaching linguistics at the Aloisianum in Gallarate, Italy."

³ See the appendix on COCOA written by Robert L. Oakman in Howard-Hill, T. H., *Literary*

Young Concordance program (BYC) later renamed and commercialized under the name WordCruncher and the TACT environment developed at the University of Toronto. This shift to the microcomputer changed the nature of our use of the tools in two ways. The scholar could now use tools whenever they wanted on a personal computer instead of having to wait for mainframe time or connect over a terminal. This meant that the humanist was no longer dependent on the paper concordance, but could use the electronic tools in his or her place of study. This change in the location of computer-assisted text-analysis along with developments in interface technology led developers away from a batch concording model towards interactive models for tool design that assumed that the scholar would have access to the tools and e-texts for personal study.

TACT is good example of this shift from batch to interactive concording. The interface was inspired partly by developments in graphical user interfaces – it has windows of sorts that are linked so that selections in one window will trigger updates in another. TACT is not meant for producing a printed concordance but for exploring the electronic text interactively through queries and windowed displays. TACT and Word Cruncher did not just automate the job of the concordancer, but changed the perspective of the user of the concordance. They are however, still resolutely textual environments that have evolved out of the concordance as a scholarly tool.

In the late 1980s and early 1990s when I worked for John Bradley in the cubicle next door and we distracted ourselves from more important tasks by imagining the ideal text-analysis environment, one possibility that intrigued us was the potential for graphical displays of information about a text. Why do text-analysis tools remain wedded to the textual concordance? Could we design tools that answered different types of questions than finding words and concording the relevant passages? We did not know at the time that Parunak in his "Prolegomena to Pictorial Concordances" had argued that distribution graphs could be thought of as an extension to the idea of the concordance. Agather we focused for reasons told elsewhere on extending the functionality of the text tool.

Hermeneutics of the Concordance

Let us pause now to consider the hermeneutical principles behind the concordance. As Willard McCarty puts it, "The early history of the concordance suggests that it was invented essentially for the same job to which we apply it today, 750 years later: to discover patterns of coherence in a text or textual corpus. ... the concordance very likely grew out of a habit of mind conditioned by a *typological* or *figural* view of the Bible, i.e. the intratextual notion that the meaning of the biblical text is derived by putting together normally disjunct passages into a *concordatia*, a concord of senses."

We can see the uses more explicitly if we look to what editors of concordances say they should be used for. In the "Introduction" to *The New Revised Standard Version*

⁵ Willard McCarty, "Introduction to Concording and Text-Analysis: History, Theory, and Methodology." Page 2.

⁴ H. Van Dyke Parunak, "Prolegomena to Pictorial Concordances."

Exhaustive Concordance to the bible the editor Bruce Metzger provides 3 uses for the concordance other than the obvious one of finding a passage. First, he says "one can make a survey of what the Bible says about any particular person or subject." Second, "A glance at information given in a concordance will show in which book or books of the Bible a given word occurs frequently, and which it occurs infrequently – or not at all." Third, "A concordance assists one in pursuing a topic throughout Scripture." He concludes with a point that includes a quote from the Bible. "The regular and intelligent use of this book will help the Bible student to appreciate that 'All scripture is inspired by God and is useful for teaching, for reproof, for correction, and for training in righteousness, so that everyone who belongs to God may be proficient, equipped for every good work' (2 Timothy 3.16-17 NSRSV)." One of the things that is interesting in this list of uses is that the uses are described in visual or navigational terms like "survey", "glance", and "pursuing".

Even more enthusiastic is Frederick Danker in his *Multipurpose Tools for Bible Study* where he says that, "The really exciting part of concordance study, though, lies in the compositional arena, where the writer's artistic and thematic competence is exhibited. ... Like finger-print powder, the concordance can disclose distinctive, latent whorls of the divine hand." Again we have a visual image of the detective finding the fingerprints of God on the text. One wonders if he had read Umberto Eco on poisoned fingerprints.

Not all discussions of the use of concordances are quite so enthusiastic. The Encyclopedia Britannica Online warns in its discussion of "Parallelism" as a form of Scriptural interpretation against the naive use of concordances. "Parallelism, the interpretation of Scripture by means of Scripture, is a corollary of the belief in the unity of Scripture. But as a hermeneutical principle it must be employed sparingly, since the unity of Scripture should be based on comprehensive exegetical study, rather than itself provide a basis. ... One naive form of parallelism is the 'concordant' method, in which it is axiomatic that a Hebrew or Greek word will always (or nearly always) have the same force wherever it occurs in the Bible, no matter who uses it."

From these passages we can extract hermeneutical principles underlying the use of the concordance and the text-analysis tools that evolved from it. I should stress that these principles apply to the use of the concordance for the interpretation of a text as opposed to its simple use for locating passages.

• First, the use of a concordance for interpreting a text presumes that there is some sort of unity to the text. This unity can be located in the artistic intentions of an author or those of a common divine inspiration. In the case of a linguistic corpus the unity comes from the criteria of selection and unity of the target linguistic community.

4

⁶ The NRSV Exhaustive Concordance. Ed. with Introduction by Bruce M. Metzger. The quotations are from page vii of the "Introduction".

⁷ Page 18 of the chapter on "Concordances" in Danker, Frederick W. Multipurpose Tools for Bible Study.

⁸ Encyclopedia Britannica Online, in the section on "biblical literature",

http://search.eb.com/bol/topic?eu=119719&sctn=14&pm=1 [Accessed May 29, 1999]

- Second the concordance allows us to create a new text that is assembled out of
 passages that agree or concord. Here is where we see the wisdom of using the word
 concordance for those passages under a single heading or those generated from a
 single query as opposed to the compilation of concordances that we call a
 concordance. The concorded hybrid provides a survey or compilation of the thoughts
 of the original author on a particular subject.
- Third, a concordance is generated according to some procedure be it a manual procedure or process implemented in software. The concordance is the result of this procedure that is more or less open to the user to recapitulate if they do not trust the results. Because we all know roughly how concordances are generated this new text's generation is rarely described explicitly in the hybrid text itself.
- Fourth and finally, to take the third point further, the concordance is generated from a query about a word or pattern. The particular concordance one looks up or generates is a text in response to a choice by the reader that is generated by the software or editor according to established procedure. This choice is what we try to develop and capture in Eye-ConTact.

The hermeneutical principles of the concordance when laid out thus raise some questions along the lines of the warning in Britannica Online.

- 1. Can we really assume that there is a unity to the text concorded such that the concordance reflects an agreement among the passages on the subject? Is the unity not partly in the question asked of the text for which the concordance is the answer?
- 2. Can passages be taken out of context to create a new collage of information with which to find an answer to a query? Even with a text by a single author, can one take passages out of context in order to traverse a topic? Is the concordance not a new text created from fragments of the old?
- 3. What is the status of this new hybrid text? Can it stand on its own as a new text? Here is where postmodern critical theory suggests that there has probably never been a text in itself that is being violated by such dismembering and recreation.
- 4. Must this hybrid text be a text? Could it not involve other media, especially measurements (quantifications) and visualizations of these?
- 5. And finally, can we imagine new procedures for generating such hybrids that reflect different types of choices by the user? Are we not limited by the type of choices allowed by a concordance?

I therefore want to suggest a different reading of what a concordance can be. Following the lines of Lucian I want to call it, and have been calling it, a hybrid text created by choices of the user from the original text. I call it a hybrid (monster) because it is authored not just by the original author, but also by the user's choices and the procedures used to generate it. It is neither afoot or ahorseback. It is neither the work of the original author nor that entirely of the user of the concordance. It's unity comes from the

intentions of both in a way that can recapitulated by others. It is one text in a larger dialogue between authors, readers, and users. Lucian had it right, we are in dialogue with Dialogue recreating monsters out of the old.

You might object that such a hybrid has no purpose. The concordance is meant to respect the original author and to provide insight into her thoughts. If we insist a concordance is a new text it loses its hermeneutical purpose. You might add that such texts by themselves are hardly interesting; divorced from their hermeneutical connection to the original they are jumbles of incoherent fragments. I have two answers to this. First, the concordance conceived as a new hybrid text generated from an original is still a text derived from the original and can still be used as a guide to the original as long as one understands how it came to be generated. All that has changed is that we acknowledge that the user's choices generated this new text. My second response is that the concordance has never been a substitute for the original; it has always been a response to a choice by the user derived from the original. Its purpose lies in its relevance to the user who looks it up or generates it. As awkward and disjointed as the new text is, it is meaningful in the context of the dialogue between the user and original.

Eye-ConTact

It may seem that I am stretching this point and that this redefinition just adds qualifications to our understanding of the obvious. Not so, the reason I have taken time to reexamine the concordance is to show how we can imagine tools that go beyond the current ones. If we step back and theorize the concordance we can see how it is but one particular type of hybrid that can be generated. The key lies in the choice of the user and the generative procedures. If we agree that the concordance is the result of particular type of choice by the reader generated by traditional procedures we can ask if there are other choices or procedures that a user would like to apply to a text. We must also, I would argue, refocus our attention away from the original text and onto the choices and procedures that co-author the hybrid text. Existing concordances and concording tools tend to hide the users role in generation in order to focus on the original. Eye-ConTact takes the a different strategy, focusing on the user as co-author of the hybrid. If I am right, the choices of the user are just as important and therefore we should imagine tools that capture these choices and procedures as part of the resulting hybrid.

From a technical point of view the issue is one of broadening the processes one can run on a text and representing those processes. Tools that evolved out of concordances are restricted in the types of questions that they let you ask of a text. Essentially you can only ask for collations of texts or distribution graphs generated by pattern searches. TACT has other tools, but these are not available in USEBASE, a clue to the limitations of the concordance style of interface. A large number of other types of programs have emerged that cannot be subsumed into the concordancer like statistical tools, visualization tools, and stylistic analysis tools. Eye-ConTact was designed to be a modular environment where the user can concatenate and run more complex procedures, including those not imagined by the authors of the environment and those that are not of the sort of "where is this pattern". It provides a visual programming interface that forces the user to explicitly lay out their choices (or program). An Eye-ConTact file or what we call a Map thus

captures not only the results of the process but the design of the process which would be the choices of the user.

Eye-ConTact

Eye-ConTact can be thought of as a working prototype of a visual programming environment for textual scholars. It is a prototype because it was designed to explore the concept, not to be shrink-wrapped and sold. For this reason it is being programmed by Patricia Monger in Visual Basic and Perl which allows us to prototype the interface and functionality quickly. It is not likely to become a distributable product because implementing it in Visual Basic means that it is slow. Should we get it right, it would have to be reimplemented. What then does it look like?

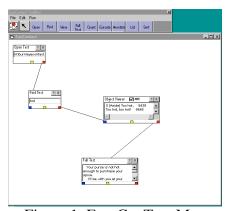


Figure 1: Eye-ConTact Map

Figure 1 shows a "Map" file in which the user had placed panels for the different processes and connected them. The user does this by selecting processes from the toolbar beneath the menu. Each panel represents a process, whether opening a text or searching one. In Figure 1 the processes start with the opening of a particular text, then flow to the searching of the text for the word "hot", so that the results can then be passed to a viewer that allows one to view a Key Word in Context Display, and finally the KWIC is linked to a full text display so that an item selected in the KWIC triggers the display of the full text at the selected point.

One of the results of building this prototype is that it has drawn our attention to problems with the initial design. In the initial design we used icons to represent the processes as in Figure 2 below. ¹⁰ We changed to panels that include actual results or important parameters in order that the Maps might better represent the choice of the user, not just the sequence of processes. In principle the Map should provide in a single image all the information you need to know about the processes set up by the user without having to

⁹ The design of the Toolbar is being reworked to allow it to include modules that the users adds to the system. The button on the far left is what you use to run the program laid out in the map.

To see screens and read an explanation of the initial design see Rockwell and Bradley, "Eye-ConTact: Towards a New Design for Research Text Tools", in particular the page at http://www.chass.utoronto.ca/epc/chwp/rockwell/ict_2.htm.

open panels. In principle you should be able to recreate the choices of the user if you had a printout of the Map. Obviously certain panels that return results can only give a peek at the results, and in some cases it is not possible to show all the necessary information without overwhelming panels, but the principle remains - if the point is to capture the choices of the user along with the hybrid text, then the Map should be as complete a guide to those choices as possible.¹¹

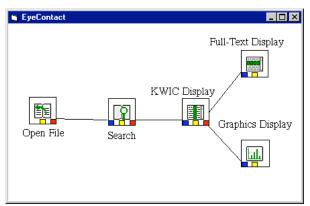
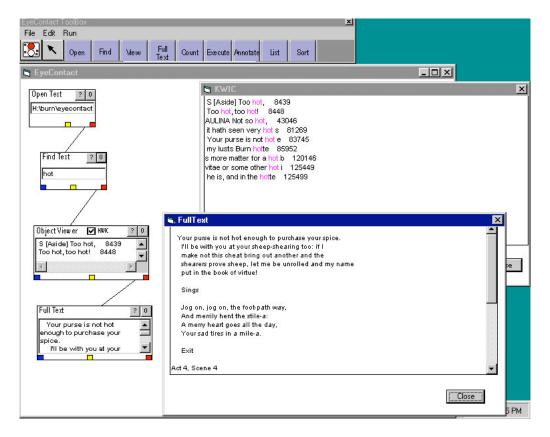


Figure 2: First Implementation

In Figure 3, you can see what some of the panels look like when they are expanded (or Opened) to display results in a more useful fashion. In the background you see the Map, while in the foreground is a KWIC display window and a full text display window.

It should be noted that there is no reason why the Map of the user's choices have to be presented visually, they could also be represented as a batch program, but we felt that a visual programming environment would be more accessible to Humanists.



To explain Eye-ConTact in terms of the study of texts - in it the scholar explicitly lays out her questions of an e-text as a visual program made up of smaller primitive processes that do clearly defined things to what is passed them. When the scholar runs the experiment on the text, there are two types of results woven into the visualization, the new hybrid text which is the result of the choices of the scholar, and the description of those choices. Eye-ConTact explicitly represents to the scholar (and anyone she cares to share it with) the hybrid that corresponds to her question and its results in the form of a visualization. Thus the software attempts to reflect the hermeneutical principle discussed above that the new text is woven from both the original and the choices of the user.

To some extent Eye-ConTact is an old idea. It is related to the old batch processing tools in that what one is describing in an Eye-ConTact Map is the flow of data from one process to another and the choices the user makes for how the primitive processes should act on their input. As such it was inspired by scientific visualization programs like Explorer which in turn are based on the data-flow model common to Unix platforms where one pipes information from one primitive program to another. In fact, in a environment like Eye-ConTact there may not actually be a flow of data from one process to another, that is simply the metaphor for the user. For example the whole e-text does not flow from the Open Text process to the Find Text process, only the name and location of the e-text flows. That said, the user is still encouraged to think about what they are designing as a flow of data through various filters.

John Bradley and others like Tom Horton have commented on the age of this model and the advantages of an object-oriented model. John Bradley is trying to persuade me that for text-analysis tools to work efficiently and for them to assist not just the study of texts, but also the preparation of texts, they should be based on a object-oriented model, and not a data-flow model. The Eye-ConTact visual programming model would still be how the user might describe certain operations they want done to a prepared text, but the actual program might actually do it by querying an object-oriented text database. Further, John has argued that this data-flow paradigm does not suit the text preparation stage where one creates larger objects out of words or passages. In Eye-ConTact that is still done outside the program by tagging the text. John has argued from watching Willard McCarty and others at work on preparing texts that the text preparation stage is an integral part of the study of a text. In a rich text preparation environment one would be able to build up a study text from the canonical one by creating hierarchies of objects like speeches in a play or clusters of linguistically related features. For this reason Eye-ConTact should be viewed as a prototype designed to work out one type of interface to what could be eventually an even richer environment.

Conclusion

To summarize let me make the following points about Eye-ConTact and the concordance.

- 1. The concordance is only one possible hybrid result from the interaction of a user and an original text. The challenge for the next generation of tools is to open the choices allowed to the user. If we think of concordances in terms of classical aesthetics they are a form of mimesis. They are representations of another text related to book reviews, interpretative essays and illustrations. They do, however have the interesting property that they are generated by procedures from choices that can be recapitulated by others in a way that other representations cannot. This is what gives them their particular rhetorical impact as works that uncover the original author's hidden fingerprints.
- 2. Because a concordance, or any other result from computer assisted text-analysis, is generated according to choices by the user, these choices should be recorded as part of the resulting hybrid. It is not enough to maintain the connection with the original text, it is also important to maintain the intentions of the user/author. This is what differentiates a scholarly tool from other text manipulators. The property of rigorous generation that gives the concordance its rhetorical character is what we want to preserve in text-analysis tools. If we are going to expand the choices offered to the representer, then we have to record those choices so that others can recapitulate them. We all know how concordances are generated which is why the process is not captured explicitly in the representation. For new types of rigorous representations to be convincing the way a concordance is we have to provide mechanisms for the readers of these representations to understand the

¹² See John T. Kirby's article "Classical Greek Origins of Western Aesthetic Theory". This article outlines Aristotle's reversal of Platonic aesthetics. Plato took visual representations to be further removed from the real forms while Aristotle argued that all works were mimesis of different sorts.

logic behind their generation. The way the hybrid text was generated is an important result, especially when non-standard procedures were used.

3. Because what we are generating are new artifacts that represent the old, there is no apriori reason why they must be concordances in the sense of written texts woven out of the old. What is important is whether these artifacts answer the concerns of the users in their dialogue. They should be interesting to their users which means that they could be graphs, hypertexts, or artifacts in other media and do not have to be simple texts. We need to expand our notion of what the legitimate output of text-analysis is. A picture generated from a text is no less a representation of it than a concordance. In fact it may be better to think of it as a visual concordance.

Bibliography

- Danker, Frederick W., *Multipurpose Tools for Bible Study*. Revised and Expanded Edition, Augsburg Fortress: Minneapolis, MN, 1993.
- Encyclopedia Britannica Online, in the section on "biblical literature", http://search.eb.com/bol/topic?eu=119719&sctn=14&pm=1 [Accessed May 29, 1999].
- Eco, Umberto, "Afterword." *The Future of the Book*. Ed. Geoffrey Nunberg, University of California Press: Berkeley, 1996. pp. 295-306.
- Howard-Hill, T. H., Literary Concordances. Pergamon: Oxford, 1979.
- Kirby, John T., "Classical Greek Origins of Western Aesthetic Theory." *Languages of Visuality*. Ed. Beate Allert, Wayne State University Press: Detroit, 1996. pp. 29-45.
- Lancashire, Ian, et al. *Using TACT with Electronic Texts*. Modern Languages Association of America: New York, 1996.
- Lucian, *Lucian: With an English Translation*. 8 vols. Trans. Austin Morris Harmon, Macmillan: New York, 1967. "The Double Indictment" is found in volume 3.
- Manguel, Alberto, "The Shape of the Book." *A History of Reading*. Alfred A. Knopf: Toronto, 1996. pp. 125-147.
- McCarty, Willard, "Handmade, Computer-assisted, and Electronic Concordances of Chaucer." *Computer-Based Chaucer Studies*, Ed. Ian Lancashire, CCH Working Papers: Toronto, 1993. pp. 49-65.
- McCarty, Willard, "Introduction to Concording and Text-Analysis: History, Theory, and Methodology." *CETH General Handbook for the Summer Seminar 1995*, Ed. Susan Hockey and Willard McCarty, CETH: Princeton, New Jersey, July, 1995.

- Preprint of Rockwell, Geoffrey, "The Visual Concordance: The Design of Eye-ConTact", *Text Technology*, vol. 10, no. 1, 2001, p. 73-86.
- *The NRSV Exhaustive Concordance*. Edited with Introduction by Bruce M. Metzger, Thomas Nelson Publishers: Nashville, Tennessee, 1991.
- Parunak, H. Van Dyke, "Prolegomena to Pictorial Concordances." *Computers and the Humanities*. 15:1, 1981. pp. 15-36.
- Raben, Joseph, "Humanities Computing 25 Years Later." *Computers and the Humanities*, 25: 6, 1991, pp. 341-350.
- Rockwell, Geoffrey, and Bradley, John, "Eye-ConTact: Towards a New Design for Research Text Tools." *Computing in the Humanities Working Papers*, A.4. February 1998. This online refereed journal is located at: http://www.chass.utoronto.ca/epc/chwp/> [Accessed July, 2000].