

Commentary, Annotation and the Virtual Humanities Lab

Vika Zafrin, Brown University

2005 NEMLA convention, March 31 - April 2, 2005

1. Introduction

Good morning! It's good to see you here so early on a Saturday. I'm here to talk about a bit of an anomaly in the humanities, a project I direct that is building electronic tools that can be used in performing quite traditional scholarship, but that go beyond, say, concordance or word collocation.

The Virtual Humanities Lab is a two-year project sponsored by the National Endowment for the Humanities. We are based at Brown University, and the project builds on the work done by Brown's Scholarly Technology Group and in the Department of Italian Studies, in collaboration with scholars elsewhere. The principal project we pursued before VHL, out of which our ideas were born, was the Decameron Web; its address and a screenshot of the Table of Contents are on page 2 of your handout.

VHL's focus is two-fold. First, we are building an online space, a web-based engine that allows us to present primary texts, for now mostly having to do with Early Modern Florence (see page 3 of the handout). In addition to simply presenting them, we are allowing scholars to annotate these texts, propose new encoding for them

(about which more in a moment), and respond to others' annotations, thereby hopefully creating and maintaining dialogue. There's a screenshot of the annotation interface on the last page of the handout; for testing the prototype we are using the full text of the *Decameron*.

The second half of VHL's efforts is going into the semantic encoding of texts. We are doing this in XML. If you're unfamiliar with semantic encoding, fear not, it's simple. eXtensible Markup Language, as its name implies, is quite flexible. It has a small set of rules and allows the encoder to specify pretty much anything she wants about a given set of data. Although we happen to be disseminating our work on the web, XML is not web-specific: it's just a way to describe information, and most of the time needs to be transformed for display.

If you have further questions about how this works, I'll be happy to answer them at the end. For the moment, let's look at an example of semantic encoding on page 4 of the handout. Here we can see names, addresses, some parts of speech and poetic structure encoded. Notably absent are instructions for the browser on how to present all these things: should the names be bold? Do we display addresses or not? Should the verbs be italicized? What does it mean when a line of a poem ends? All of these can be specified using other languages, the most famous among them HTML; for the moment, we are interested in isolating and naming semantic structures.

Semantic encoding can function as a scholarly argument. The second example (on page 5) posits that from line 3 to line 4 Boccaccio was talking about geoastronomy. There are many places in the 700-page text where he seems to touch upon Earth's relationship with the rest of the cosmos; we are encoding all of them. This encoding is, of course, disputable. Someone may come along and say that no, here Boccaccio was just making an offhand metaphorical reference that doesn't actually have anything to do with astronomical phenomena. More often, people might point out places in the text that we *haven't* encoded as relevant to geoastronomy, but should. By bits and pieces, a bird's-eye view of the text is constructed, which may reveal larger structures in it, or in a corpus of texts, that are difficult to see with the naked eye.

Encoding once does not preclude a division of the same text into a *different* set of semantic parts, with another purpose or from another angle, or in response to an argument made through previous encoding. Each variant model contributes to a deeper understanding of the material at hand.

2. Collaboration

For the remainder of the presentation, I will make some generalizations about humanities scholars and scholarship. Please be assured that they are self-conscious generalizations, hypotheses

still being tested out. If the claims seem too categorical, it is due to time constraints and not a lack of unanswered questions.

That said, humanists seem to have a love-hate relationship with collaboration. On one hand, the very purpose of humanities research and writing (along with preserving and propagating knowledge) is communication. If you want to make it in your chosen field, you must make yourself known to other scholars by presenting at conferences, publishing articles and books, and keeping abreast of new developments. But this is one-way communication. What really seems to charge us is when an audience shows interest in our work by engaging with us, asking questions, proposing alternative opinions, offering new information. Conversations over coffee, electronic correspondence and publications that directly respond to previous publications often accomplish more, and more easily, than a lone scholar can do even in the best of libraries. Communication is rewarding emotionally, as well: it acts as direct confirmation that we are not working in a vacuum.

And yet, when it comes to actual work *projects*, unless there is a book to co-edit humanists seem to prefer to work alone. Conference papers more often than not have a single author; likewise with publications. Bouncing around ideas in coffee shops certainly carries a rush, perhaps a productivity spike, but institutions do not yet know how to recognize work done by more than one person, where to draw the boundaries, how much credit to give. As a result, promotion and

tenure reviews *tend* to look at the work on which only one scholar's name appears. This despite the fact that informal collaboration happens all the time, as people email their colleagues bits of information, pass them references to articles over lunch and so on -- such that, in reality, few publications are *actually* entirely single-authored.

In his *A Social History of Knowledge from Gutenberg to Diderot*, Peter Burke presents much historical evidence of possibilities opened by direct contact that may not have come to be at all without it. My favorite is the series of coincidences that led to a meeting between a 17th-century Italian sinologist and a Dutch arabist, each of whom happened to possess rather obscure knowledge that the other found useful! [I believe it had something to do with putting religious texts into their cultural context.] They could conceivably have advanced their studies on their own, but even if that happened at all, it would have taken a lot more time.

Lately, we have again placed increasing value on direct contact. Humanists have started blogging, for example. They post thoughts about their current research, the state of academe in general, and other topics of interest on the web using software that encourages comments and therefore conversation. Some of these weblogs have taken off. The Invisible Adjunct evolved from an anonymous professor's perspective on academe into something of a community, which was eventually written up in the Chronicle of Higher Education.

Grand Text Auto is a popular collective weblog with five authors, dedicated to computer art and literature as well as games. There are blogs devoted to teaching, women in technology, and a slew of individual blogs written by professors of English, history, political science, you name it.

Weblogs are not, of course, the only collaboration tools out there. One example in particular stands out. For the past several years, six universities all across Canada have been developing a project called TAPoR – Text Analysis Portal for Research. They have funding from the Canadian government, and all of their resources are freely available to researchers in Canada and abroad. While their tools are still in development, they're already pretty exciting. For example: if you have an account (test account for now), you can upload texts from anywhere on the internet, including your own computer. Then you can use tools to do all sorts of things to your text: visualize it graphically according to word frequency, find word co-occurrences and collocations, list all of the encoding elements you've used if that's relevant, and many others. And here's the really exciting part: eventually, you'll be able to do this *across* texts, incorporate tools developed and housed by other people at other universities, and *share* your work with your colleagues, with your students, with your spouse. Group accounts should also be possible, so that multiple people work on texts together. You get the idea.

An open-access approach is also a potential downside of

collaboration. Scholar sceptics are (rightly) wary of endeavors that allow everyone, regardless of qualification, to present their opinions on an equal standing. Objectivity in deciding who gets what level of access is a tricky and separate issue. That aside, the VHL toolset is aimed specifically at textual analysis and criticism, and will implement different levels of access and offer reasonable levels of peer review. We have drawn up a tentative user hierarchy (page 6 of the handout), and will modify it as needed in the course of the project's development.

For all the access constraints, you'll notice that even a person who stumbles upon the project out of nowhere will still be able to contribute. The point is not to create a closed system; on the contrary, we would like to create an open one that is able to place appropriate weight upon input from different sources.

3. Participation

The texts we have chosen for the first round were no small part of the inspiration for this project in the first place. Both Boccaccio's *Esposizioni* and Villani's *Cronica* address their audience directly, peers and laymen alike, explicitly and implicitly. Although half a millennium later we treat them as more or less stand-alone books, the *Esposizioni* in particular was not a book at all but a series of lectures, and was part of an ongoing debate among the literati concerning the Divine Comedy. Villani was writing his chronicle more or less alone, but his

book lends itself to an exploration of family-based Florentine politics and the main themes of the day-to-day life (such as, for example, different types of violence, on the part of both the authorities and the populace). Pico's theses, once they are online, can be discussed and debated live, in the way he himself seemed to want when he requested a defense before a commission of cardinals and Catholic theologians in 1486.

Eventually we hope to open up VHL for new texts. For the two-year duration of the project, however, we'd like to concentrate on getting the annotation and discussion systems up and running, and making sure they are stable and easily accessible.

4. Adaptation

So what about adaptation? Well, we at the VHL feel that we're participating in a process begun by others: that of adapting the humanities for the electronic era. Sounds grandiose, but humanities computing has now existed as a field of inquiry for over fifty years, and is quickly gaining ground. There is no question about the usefulness of computing and the network for both research and teaching. Beyond its quotidian usefulness, examining texts using computation and semantic encoding provides fodder for new ideas. Like Henry Ford, who tinkered with machines and modeled them in order to come up with the first more-or-less general-use automobile, we want to tinker with our primary resources, go in with a hunch and

see what happens.

We are also adapting scholarly editing from a practice that produces a *fait accompli* to an ongoing process. Electronic scholarly editions have been gaining credibility, and especially when it comes to old manuscript texts, or for that matter paintings, sculpture and music, it seems that the electronic medium is no further removed from the primary source than a mass-produced codex. This is not to say that electronic editions will, or should, replace paper ones; the two are different animals and one should not substitute for the other.

Finally, we are adapting the scholarly practice of annotation for collaborating purposes. Annotation has been critically useful for both the annotators and the readers, and conversation leads to more active and exciting thinking; so we're bringing the two together in a community-building effort. We hope that humanities scholars will find it a useful tool and an inspiration. If any of you would like to collaborate, we'll be thrilled; if you'd like to critique, please do: we want to make something people will actually use, and welcome constructive feedback.