Stalking the Paratext:
Speculations on Hypertext Links as a Second Order Text

Francisco J. Ricardo

Harvard Graduate School of Education
201 Longfellow Hall
1 Appian Way
Cambridge, MA 02138, USA
Tel: 1-617-496-2729
E-mail: ricardfr@hugse1.harvard.edu

ABSTRACT

In the popular conception of hypertext as nonlinear writing, primary emphasis typically falls on the construction, character, and quantity of constituent lexias that comprise any given hypertext. This paper, however, will focus on what the text would reveal if an ordered collection were made of the links emerging from the main (first order) text. Such a collection, as a second order text or parallel text, which I propose to call the *paratext*, comprises the layer-world of links, of intertextual referents that could be subjected to cluster analyses that reveal aspects of cohesion, breadth, and other speculative characteristics of the first order text.

KEYWORDS: Hypertext, intertextuality, link semantics, grammatology, paratext, rhetoric.

INTRODUCTION

This paper discusses the theme of augmenting possible understanding of a hypertext by applying an interpretive analysis of its links. The first part of the paper describes the concept and framework of the approach in critical terms. The second part describes aspects of a prototype application approach being developed within this framework in conceptual and algorithmic terms.

For some time now, the importance and implications of relating hypertext to intertextuality by way of links has been broadly described in various disciplines. In literary theory, the need for intertextual links as a way to augment the reader's interpretation was originally discussed by Kristeva, and later expanded upon by Derrida, Barthes, Greimas, among others. Likewise, in more applied domains like computer science, hypertext and its notion of links is seen as a tool that provides flexible access to information by incorporating the notions of navigation, annotation, and tailored presentation of textual material [9].

There are traditions in the lifeworld culture, in formal education, scholarship, and some branches of literary theory that approach texts conservatively, led by the overriding assumption that the author invariably desires a single interpretation, that the conclusion is obvious if properly analyzed, and that the aim of reading is to "get" the author's point. It is easy in a conference on leading edge grammatology to dismiss these views, except that they are believed by much of society outside the academy. The prevalence of these views, in my opinion, is central both to the resistance to, or misunderstanding of, hypertext by reasonably intelligent people.

And so, this paper will explore two perspectives. The first is an argument that acknowledges textuality as the historical antecedent of intertextuality, embracing the popular conception of hypertext as an aberration of exposition; the second is an implementation approach. These strands revolve around the notion of the link as an interpretive mechanism.

Argument

In excessively superficial speed, I would like to outline three epochs of writing, so that the position of the hypertext link can be seen as a bridge to a new grammatology. Each age has had the use of a symbolic mechanism to define its primary signifying practice. I would like to argue that for each epoch, these are the token glyph, the sequential narrative, and the hypertext link, respectively. To commemorate the third epoch of writing, the hypertext link will be made to carry its own signification, much as narrative has become its own kind of study today. So, in an attempt to add to that scholarly pursuit, I will propose the notion of a paratext, a dimension of signification that begins within the text, but might systematically be shown to spawn its own narrative depictions.

Epoch One: Token and Archaic Writing

Perhaps the earliest known writing comes from the Sumerians of Mesopotamia. The Pergamon Museum in Berlin possesses evidence from one Sumerian city-state, Uruk, from which were recovered thousands of systematic texts created six thousand years ago. This writing has been traced without significant modification to a Neolithic precursor created eleven thousand years ago, and the real question of how the intervening five thousand years of writing could change little enough to be recognized has been answered with the postulation that this interim period had no need of a complex evolution, since the practice of written signification was limited to the simple record-keeping of goods and trades [19].

All early forms of writing are either pictographic or ideographic. Pictographic glyphs resemble visually what they represent in the physical world; an ox looks like an ox. An ideograph, on the other hand, is a signifier from which the signified cannot be deduced visually. It may instead, for instance, be encoded into a phonetic symbol system, in which case the glyph or word decodes to the *sound* of the signified, rather than its visual appearance. But both systems are essentially forms for annotation of single entities.

This type of writing connotes a symmetrical, semiotic mapping: a thing in the world is signified by a glyph, which in turn signifies and evokes it. The use of writing implies its learning, and the idea of literacy in any culture or historical moment is the skill of generating, at any time, one side of the signification, having only been given the other. Literacy, like remedial reading, is atomic: it specializes primarily on meaning up to the level of an utterance. In this *token-oriented* emphasis, literacy is part of a larger archaeological view applied by scholars of ancient and extinct cultures and their writing systems. For them, the Holy Grail is the archaeological semiotic of any Rosetta stone found decipherable for linear scripts and pictographic writing systems.

Epoch Two: Sequence and Modern Writing

If another Rosetta stone *were* found after the demise of our own culture, it would probably not be very useful as a device for successful interpretation of our texts. That is because our own writing has evolved, and the locus of signification in a text is no longer at the level of any individual object, symbol, or utterance. What has happened in just a few millennia? Writing, which began as the technology of token annotation, somehow became infused with the problems of sequence. Modern writing is no longer in the practice of signifying singular things like bread, sheep, and oxen. In serving the function of simple accounting and trade transactions, the focus of archaic writing was the *commodity*. But subsequent sacred writings were not about the denotation of material things at a particular point in time, but rather the capture of thematic strands of human *experience* over and across time. We can in this way distinguish Egyptian hieroglyphic writings from those of Classical Greece. For, even in the afterworld, the Pharaohs were inventorying their possessions, there is no evidence of introspection in the hieroglyphs. By contrast, the Greeks were from their first days exploring questions of being, action, and destiny: ontos, ethos, and telos.

To tackle questions of human experience such as these, it became necessary to tell stories as an important device for diverting the reader from applying the literal, tokenized, commodity-oriented kind of interpretation that had been the purpose of writing for centuries. When the element of time enters the text in this way, we move from primitive exposition to extended narrative, and it is no longer possible to locate a center of signification in the text. The loss of obvious "signified" in the story marks the end of writing as a mechanism exclusively for depicting commodities, and more for depicting processes. The signification of entities in the text is the analytic domain of semiotics, but that of processes belongs to hermeneutics.

The genealogy of narrative emerges with the fact that, in order for any human experience or process to be revealed, it must be revealed *over time*, and because of this, there is a fundamental chronological essence to narrative. Narrative begins when signification is defined more in terms of sequences than commodities. In fact, tradition reveals the controversies over the best kind of interpretation for old texts. In the literal view, Moses really parted the waters. In the semiotic view, the waters represent the unrefined emotions, which are conquered by Moses, the human sign of the moral nature of Law, and progress can ensue. In the hermeneutic view, the waters are not the central symbol, and neither is Moses; rather, the process of parting or clearing the way is the crux, and it makes sense only in an extended, chronological context of departure, straying, disobedience, marginalization, and finally persecution. Each of these themes is a link to the great story of a people, and without them, the story lacks thematic depth and fails in its objective of bringing the reader to the point of becoming convinced, reflective, and ultimately transformed.

In the span of time covered here, writing itself has been transformed by the introduction of chronology, supported by sequence, of ideas, of thoughts, of depictions. The sequential nature of writing, then, is here seen as the opportunity to enrich signification by broadening the context of what is depicted with supporting themes, episodes, and examples. All of these "stand in" for the reader's interpretation until the last word has been read, and at that point what, if anything, is taken from the reading is the aggregate of intertwined descriptions understood either as a plurality of points or as a single symbolic experience with many prophetic preambles. The moment we choose to believe that a conclusion was predicted by illustrations, instances, or patterns, we are transformed by the text into the cosmology of its thematic causalities. Where there are scenarios, examples, or anecdotal incidentals, there is narrative, and, inevitably, the brand of persuasion we call rhetoric. But in the second epoch, subtleties in sequence mean that signification and texts can no longer be "decoded", as was archaic writing.

Epoch Three: Writing After Sequence

Once sequence is abandoned in favor of alternatives, we begin critical reading. But once the same happens at the level of *production*, we speak of links, of hyper-writing, and of an end to persuasion. Nothing will turn back the clock once irrefutable research, theory, and application begins to prove that linear sequence is less desirable or intuitive than was thought. This is what hypertext proves, even if it is not really what Aristotle had in mind when he devised the forms of rhetorical structure that have become prevalent today. For thousands of years, our Western tradition of discourse has made it clear that there can be no latitude for argumentative uncertainty in the exposition. We have come to believe that if our writing is open to interpretation, if it can be rendered in manifold ways, then it is not probably precise enough, and perhaps incorrectly expressed. For our venerable tradition rests on the classical *technè rhétorikè*, whose five operations of discourse involve volitional acts minutely controlled by the author, namely *inventio*, constructing the assertion; *dispositio*, constructing the order; *elocutio*, constructing the style; *actio*, constructing the delivery; and *memoria*, constructing the body. It is interesting that, of these, *memoria*, the last, is the most potentially intertextual form, as Barthes has pointed out [4]. Yet, it disappeared, as did *actio*, when spoken discourse was displaced by writing as the primary rhetorical medium, leaving argument fully in the control of the artificial form. There is no way around the inevitability of conclusions built into the rhetorical structure of a traditional text. By curtailing any competing interpretive links that might take the observer to other associations— and other conclusions—rhetorical text is not merely argument; it is coercion. For Landow, going from hypertext back to the linear world of print brings frustrations attributable to "repeated recognitions that effective argument requires closing off connections and abandoning lines of investigation that hypertextuality would have made available [15]. Classical, linear rhetoric is the archetype of discourse embedded within the collective consciousness of the rational Western mind, and hypertext stands in non-negligible opposition to it.

Accordingly, modern printed text benefits from a long tradition, and through time, it has matured and become rather conservative in style and substance compared to other expressive modalities. Today, nothing solemn or dignified is ever fashioned without being represented in some textual appearance. No modern person, for instance, would recognize a *contract* as valid without first seeing it on paper. Even the authority of the person is not valid until it is textualized into a *signature*. Text is the stage on which meaning performs in a predictable way for history. In film, by contrast, meaning is always overloaded and open to many interpretations, so, despite the ubiquity of that medium, it has not achieved the same distinction. It does not want immortality or legalistic credibility if the price to pay is dogmatic immutability of representation. Film, like other communications media, has no rhetoric that moved from the flexible to the fixed, as has classical rhetoric in progressing from textual orality to textual literacy.

Maximal control of interpretation is the exclusive domain of text and no other medium, and this is apparently the historical rationale for its staying power.

It is clear, then, that sequential text has attained an exalted position, a privileged status, among all other forms of communication. It certainly would resist attempts to reconfigure the trajectory of its rhetorical history and future. To a modern reader, hypertext is different and confusing, it presents a threat to the established approach of Aristotelian comprehension, even if such comprehension has today become reduced to many trite and overused forms.

The definition of hypertext itself is currently in some state of reconfiguration, and is divided into various competing definitions, including cybertext, ergodic text, and other intriguing terms [1]. But to the average reader, the fundamental problem with hypertext is not that it argues differently, but that it reads differently; it is, after all, nonsequential writing [18]. And this most succinct of definitions is for many people an unacceptable oxymoron.

Let us also remember that nonsequential writing, to use Ted Nelson's phrase, describes the *process*. What emerges from nonsequential writing is the *product* of hypertext, and that is where our troubles begin, because we cannot say what this product is. It is not a book, or if you like, it is not the same book next time around. It is a written text that is divided and then, by means of links, made whole again, in some new form. Of course, we could also say that the word "form" is not the most accurate term for something that is constantly being interpreted in different ways and through different paths. We could say that hypertext is not forming text, but reforming it, which reveals why hypertext has many enemies who don't think text is in need of any reform.

More pragmatically, the problems of delimiting form in this medium imply that hypertext dissolves the difference between the responsibility of the reader and that of the text: to the degree that paths, trails, links, and webs are its essence [13], hypertext does on screen some of what the print reader does mentally. Joining visibly in one interface what normally happens in two very separate worlds, a mental and a digital one, means that hypertext itself executes part of the reading with the reader—a recursive phenomenon. This is perhaps why Joyce has persuasively argued that hypertext, "at least on an impressionistic level, has the potential to foreground elements of experience that classical physics, epistemology, and maybe even the technology of the 'book' erase from the scene." [6] The primary point here is that hypertext is not entirely an object, but is also partly a subject. Since it can't make itself into a complete object, its form is vexing; when we read, it is not entirely separate from us.

Reading is *traveling*; there is always a path traversed for the duration of any reading; it is why narrative is historical in flavor. Conjointly, we could argue that *all*

comprehension—textual and otherwise—is temporally determined; for every present in the reading, there is always a past, whether or not it is in the particular text. In an immediate sense this past is, of course, the context of the topic; in a more global sense, this past, as Kristeva explains [14], is the culture of the signifying practice, which informs what is not explicitly written because it is customarily understood. The understanding of common exclusions is a problem for all discourse communities, and largely defines membership as the set of those who can understand the omissions. Similarly, hypertext is itself a discourse community whose members are readers who accept rhetorical absences with the faith that in the end, meaning will coalesce. But the conventional reader thinks differently, and craves rhetorical linearity in the text, because it adds to pre-established context with each new moment of reading. The hypertext reader expects that context will emerge after the reading, but for the traditional reader the text is what comes after the context. The progressive hypertext reader is aware that there is a difference between point and story, but for the traditional reader there is no difference between context and what comes before the text. One can be more easily persuaded if one cannot differentiate context from *pretext*, since both can precede and influence the reading. Context is the unnoticed expectation of the reader; pretext is the unnoticed purpose of the writer, and hypertext separates both by churning context so that pretext cannot take hold.

Clearly, then, for the traditional reader, every hypertext must be sensitive to how context is progressively fashioned from the passages in its corpus. Consequently, after the hypertext reading, at least one of three typical outcomes is that the reader may have derived some compressed learning, may have perceived numerous complementary perspectives on a topic, or may have experienced a general sense of incoherent confusion, no doubt exacerbated by the plurality of links and pathways [7]. The outcome in each case is caused by the correlation of two determinants: the reader's search for traditional rhetorical linearity that is felt as thematic *continuity* on one hand, and the degree to which the text and links argue separately and together a case for different possible thematic continuities. The hypertext reading experience, then, is the result of a compromise between one selected thematic linearity and the universe of possible linearities in the hypertext.

This first determinant of the hypertext reading—the experience of continuity—is the product of that classical, traditional, *rhetorical* approach of putting the text at the center of meaning, which has been inculcated throughout our education, and is thus what one would expect from a traditional orientation. But it also characterizes the hypertext's links as loose agents without any plan other than to serve the reader with a choice at some particular point. This loss of continuity is why some people not familiar with hypertext express repulsion to it. I think that their perspective, which I'll call the analytic, is problematic; centering on the *textual* rather than the connective aspect of hypertext gives little independent

thought to the species of the link as its own phenomenon. Aristotle is alive and well today, it seems, and he doesn't like hypertext. It won't let the author define the debate using his predefined rhetorical framework. The analytic perspective considers the truth-value of any statement to be provable exclusively by the propositions in it; there is no exterior reference or larger, fuzzy context to enter or influence the conclusion. Analytic philosophy today is the most salient lineage left by Aristotle to the modern mind. Too bad that hypertext is framed in more continental terms, for hypertext isn't just a signifying practice, it's also a critical theory of reading, and for critical theory there is no escaping the role of larger context—and the fuzzier the better.

The analytic view favors propositions, which means static objects and fixed representation. This favors signification, but deprecates processes. That is, analytic philosophers are uncomfortable with things through which other things themselves become signified. This view is naturally uncomfortable with hypertext since, through links, text is transfigured from message to medium. Just as the many transitory digressions of a conversation gradually shape the listener's progressive understanding, links likewise cannot help but bring about subtle transformations in the emerging interpretation peculiar to the individual reader. Links deny the closed validity of an analytic, Aristotelian system, but in turn offer an extremely hermeneutic reading device.

The second determinant of the hypertext reading—the degree to which the text and links argue separately and together a case for different possible continuities—is evidence that links have a hand in any conclusive interpretation, because, in transposing, rearranging, and dislodging the text, links disavow every chapter in the annals of *technè rhétorikè* that aims to construct that assertion, that order, that style, that delivery, and that body which have become so familiar to our understanding of text. Discarding the traditional goals of the analytic reading becomes the price of entering the hypertext, where links as much as words are the primary enzyme of semantic construction.

Both determinants of the hypertext reading revolve around the challenge of ascertaining the principal question motivating the text. Gadamer argues that "the most important thing is the question that the text puts to us…The voice that speaks to us from the past—whether text, work, trace—itself poses a question and places our meaning in openness." [11] Likewise, the problem of reading any hypertext is the problem of identifying, not the question, but the *world* that it puts to us; in this sense, hypertext is a hermeneutic principle. Any hypertext may contain and advocate multiple continuities, but there is still a connective question behind it which the reader must come to understand as a condition of understanding the reading itself. Through it, we can foster some understanding of what we're doing for the Aristotelian, analytic, and traditionally oriented readers. Hypertext is exciting, but their misunderstandings occasionally relate to iconoclastic

claims (propagated by our own community) for hypertext as the liberating juggernaut that, by means of links and spatialized webs has rendered interpretation obsolete, displaced the wholeness of the work, and even assassinated the author.

What the traditional reader has not been given is a sense that hypertext is about plurality and comprehensiveness *as well a*s singularity and focus. Sequential text can be understood by the semiotic philosophy of Greimas, but hypertext is probably better explicated by the process philosophy of Whitehead. For it is not that any interpretation is outmoded, but that hypertexts are in effect *communities* of interpretations. It is not that the wholeness of the work is displaced or even destroyed by the presence of the link; it is rather that linked documents have a more comprehensive wholeness than can be embraced in a single reading or summarization. And it is not that the author has been murdered, but rather that the creator is less important than the creation, and I don't see that as impoverishing the reader. My view is that the author can always fend for himself; it's the reader we must protect from extinction.

I have been leading up to the more or less widely recognized distinction between *uni-sequential* text— represented by the spirit of Aristotelian rhetoric, analytical philosophy, and even semiotics—and hypertext, better understood by continental and critical theory. I won't recapitulate the seminal work of Landow or Joyce, but where the elementary distinction begins is in seeing that hypertext contains explicit links and text does not. What is important is that this distinction permits link-related research unique to hypertext, involving implications of navigation and disorientation [10]; distribution of documents as a single body of interrelated text for what can be called social reading [3]; strategies of content and structure [17]; library and archival problematics [16]; access and database systematics [20]; and other classes of problems that obtain when text is divided, distributed and again reintegrated by mechanisms of functional interrelationships.

Emphasizing the reintegration of text by such relationships is what distinguishes the engineering philosophy of hypermedia systems. That is, Notecards, Microcosm, and Intermedia, to name three imminent hyper-environments, can be compared by the way in which they bring texts into function. The link features in these systems become more than an operational element; they have become part of the semantics of the text. And when this happens, we embark on a new and very post-rhetorical, neo-Aristotelian tradition.

Those drawn into this interesting turn of events might in fact explore a central question: what are the relationships between hypertextual linking and subsequent interpretation? More simply put, what does linking do for interpretation? Putting it even more directly, if links could talk, what would they say about their texts?

One embryonic way to explore this question is to treat links separately from the lexias they interconnect, and expand their descriptive power to the point where, by themselves, they become a reading. And it is toward exploring this possibility that I now turn.

Implementation

The previous questions call for a need to interrogate links.

We designed a piece of software that scans a hypertext and collects up all the links that emerge from it. In HTML, this is indicated by the pattern of HREF tags

```
<A HREF=destination>referent</A>
```

The referent is the link name or description of the destination node. Then, by the procedure defined below, the program will process the destination toward which each of these links points and determines the type of relationship that each link asserts in bridging the two texts. The problematic term here is the word "relationship" and I should clarify it.

There are classification structures for drawing relationships between constituents of a text. Some, like Trigg's link taxonomy [21], have been quite comprehensive. Others limit their classifications to more specific constituents of a text, such as its structural sections (announced by headings, for example) [8], or by less structured lexical scanning of the work [12, 5]. All such work has been important because the determination of relationships within and among texts announces the opportunity for interlinking.

The current work of this paper does not build links, nor does it type links in a system according to relationships in the strict sense. The determination of link types, as well as the description of text based on the link characteristics comprise important problems, but the overriding question of the current investigation is what a link might say about its text, not what relationship obtains between any two or more texts. The latter distinction is more text-centered, whereas the former is a strictly link-centered question. That is, when we categorize a link by virtue of its relation to any lexias, we are making a connective statement about the lexias. To be sure, the link's name is the name of the relation that one text gives the other: support, data, argument, problem-posing, commentary, etc. But the link, by virtue of being named something of interest only to the two interrelated texts, loses some autonomy.

Imagine that the links, identified by the referent text given in the HREF tags of an HTML page, were collected up, and that from this list, each link-relation itself were then mapped onto a super-conceptual category space exactly like what we imagine when we think of Roget's Thesaurus. That is, this space—a large text document by itself—is a terrain of categories arranged in such a way that concepts closely similar to each other are adjacent neighbors in the space, but those with few characteristics in common are far apart from each other. Groupings of this type have been done before [12]. The relevance groupings create local

neighborhoods to signify the closeness of their referents, but the overall semantic space is frequently a closed and rather specialized knowledge database suited for specific tasks.

In order to determine how a link "understands" its surrounding text, the interpretive framework for that understanding doesn't have to be a specialized knowledge base; a broad semantic space, like a thesaurus would also do. An ordered corpus such as this could reveal aspects of the link not as obvious in specialized systems. These link aspects could be understood as a set of increasingly elusive challenges for the hypertext researcher.

For example, one relatively manageable advance of hypertext research is the automatic ability to scan two texts and establish cardinal points of similarity between them for the purpose of associating them with a new link. Bernstein's exposition of a link apprentice gives an excellent overview of the problems involved here [5]. So at this first point, the link only exists to connect ostensibly relevant but non-sequential content.

Something more difficult but still not impossible is the task of making a link automatically explain the relation it can convey between any two texts. James Allan's work on automatic hypertext link typing is among the most informative in this direction [2]. So at this point, the touchstone among two lexias is not only found, but the inserted link additionally expresses *some* of the relation between them.

Something even more intricate, however, is the ability to determine the *degree of abstraction* that a link conveys in relation to its texts. It would be useful in a hypertext on automobiles, for instance, to automatically classify links by degree of specificity on the topic, so that links connecting to informative nodes about travel come first, and then the reader can link to those explicating horsepower, suspension, and transmission gear sets. This can be accomplished only by relating links through the widest possible conceptual category space.

The source for the conceptual space in these experiments was an electronic version of the 1911 edition of Roget's Thesaurus published by the Crowell company and prepared by MYCRA, Inc. While several significant thesaurus preparation projects have been underway over the years, the reasons for this choice were simple: this file is in the public domain; it is more complete than some more recent editions; and each synonym is collected into a sequence called a main entry. Fortunately, each main entry is also numbered and bears a pound sign "#" in front of the number to facilitate computerized search. Finally, and most importantly, all main entries are arranged by semantic category rather than alphabetically. This enables the automatic determination of semantic similitude or closeness of relation between any two or more words by comparing the proximity of the conceptual entry numbers into which they fall. This thesaurus, in other words, has built-in

support for relevance groupings. Moreover, the actual number of a word's main entry is an indicator of its abstractness. The earlier the entry number, the more abstract the concept. From entry number 1, denominated as Existence, the most fundamental ontological category, the thesaurus traverses one thousand entries covering increasing concrete concepts: Form, Motion, Matter, Action, Relations, Institutions and so on. It is true that any taxonomy is a mechanism full of descriptive arbitrariness, and in the current case, other categories could have been argued, to be sure. This particular corpus, for instance, is partial to religious and moral sentiments and connotations, which comprise twenty percent of its categories, while minimizing scientific or mechanical terms and functions. But for narrative analysis, such partiality is convenient and probably even favorable.

This thesaurus works like any other: any concept could be ostensibly located residing in some specific place, under some category. Consequently, if each link from a hypertext were to enter this space, it could then be situated in its categorically relevant location. We could repeat this process of injecting each link relation from the hypertext into its corresponding category into this conceptual topography. Once every link referent has found its conceptual position in the map, the final step would be to produce the correlation of proximities that obtain between our collected intertextual relations when they are mapped out over the terrain of this thesaurus. For in this space, we might see not only that links made around certain similar concepts would cluster together in some conceptual sector of the thesaurus, but additionally that there might in fact be various clusters of conceptual kinship that would be favored by the text, because a greater number of links populate that given region.

What would this say about the text? When link referents cluster in a semantic neighborhood, they provide a consistent reference to a set or family of concepts, and thereby show that hypertext links can occasionally sustain separate meanings and communicative intentions of their own, hence the idea of a separate, alternative, and parallel text, or *paratext*. Though cohesive, the paratext may not necessarily be anticipated or intended by the author, so it becomes important to wonder its degree of relevance or proximity to the larger, foundational lexias of the hypertext.

Approach
The origins of a paratext begin with a traditional hypertext schema. Most simply outlined, the schema would comprise lexias interconnected by links in any configuration. The greater the number of links, the better, since the system reads links rather than lexias. This is a fundamental assumption of the current system, and in this, it is different from link construction systems and apprentices, which scan text and then build links from their processing of the text. To explore the paratext, the links must already have been authored into the hypertext.
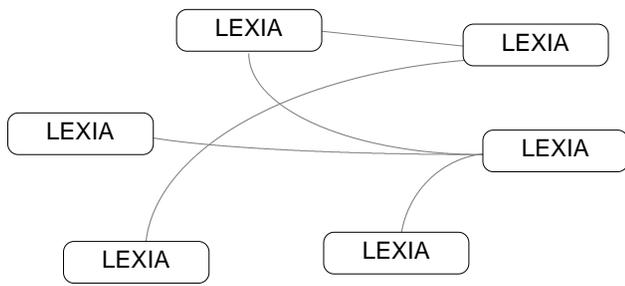
Figure 1: A Traditional Hypertext Schema

Of course, this could describe a simple web site, so that the pages branch out from the hypertext links like this:
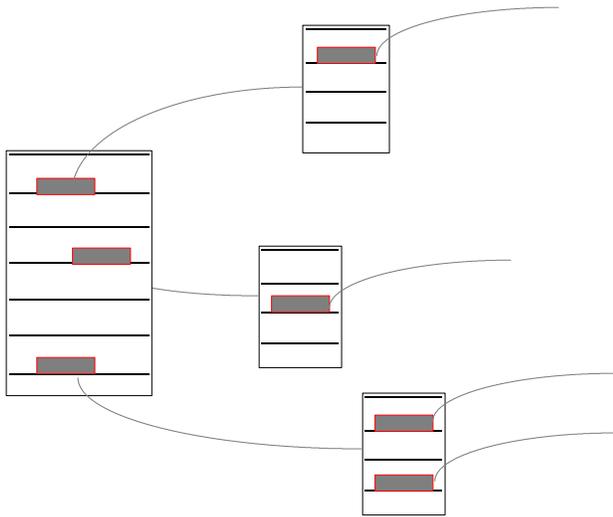
**Figure 2: The Hypertext of a Web Site**

Next we construct an inventory of the link referents, the words used as descriptors of the destination toward which HREF tags would take the reader. The descriptive accuracy of a referent to its destination node is quite variable and depends on authorial conventions of embedding descriptive information in the link's HREF tags.

It should be said that, in some cases, a link's referent might contain a very precise description of the node, in other cases, the referent can be wholly uninformative. The least useful such referents are those that state neither the direction nor what might found at the link's destination. The destination may in fact be well described, but unfortunately only in text that precedes and is outside of the HREF link tag. For instance, in a page that states, "If you're interested in a detailed analysis of how Titian's late style led to the emancipation of painting from the ideologically religious repression of the Vatican, click here" the link anchor would be the last word.

A less nondescript, but equally problematic case is that of link referents that provide generic direction but no understanding of the destination. Examples here are the ubiquitous "Go Back," "Top of Page," and the like. These utility links are convenient, but sometimes not as crucial as a hypertext author might think. Either way, their presence surely does nothing for any automatic scan or assessment of links in the corpus.

The descriptive depth of a link, as laid out in the HREF tag, is (for me at least) a measure of authorial maturity in that it reveals how well the author understands the medium of hypertext. In many cases, a web author uses links more for navigational convenience than as an active means for fostering depth and diversity of interpretation. In such cases, a paratextual analysis reveals little semantic coalescence and correspondingly little evidence of links as a stimulus to the reading.

This semantic coalescence which is the crux of a paratext is derived from the inventory of the link referents by a simple numerical procedure.

The link name used in each link on each page.
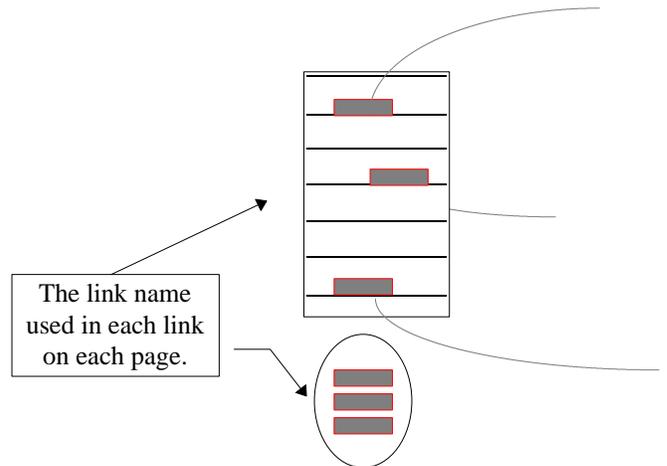
Figure 3: Inventory of Link Referents from a Web Page

For each link referent, we search the thesaurus, and extract not only the main categorical terms (or main entries) which capture it, but also their category numbers. Now each link obtains a semantic space of categories by itself, which might resemble something like a comma-delimited string of thesaurus concept numbers where the referents were found:

| 313, 515, 546, 917 |
| 110, 113, 115, 120 |
| 19, 45, 920 |
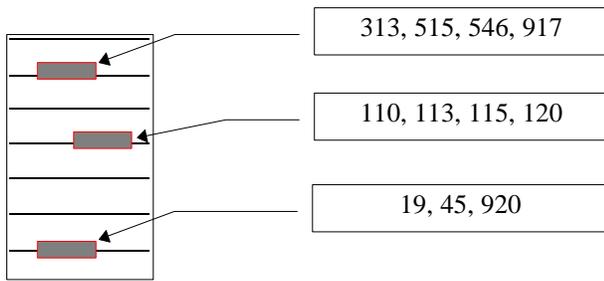
Figure 4: The Semantic Category Space of Each Hypertext Link Referent

Dimensions of Paratextual Inference

What is derived after the application of these audit mechanisms? There are three classes of inference that the links can justify, once mapped from a conceptual space. Each is the result of some regressive analysis of the numbers in each link's semantic category space. As a measure of interpretation, a paratext reveals three axes of description: a referent's conceptual abstraction, its linguistic applications, and its semantic connotations. These are determined by measuring three forms of the referent's location in the thesaurus: we determine the referent's degree of conceptual abstraction from its position in the thesaurus; we determine its linguistic applications from the quantity of its instances, and we determine its semantic connotations from the range of its appearances in the thesaurus.

Axis 1: Establishing the referent's abstraction from concept position

Earlier it was mentioned that a word's concept number is an indicator of its abstractness. The earlier the entry number, the more abstract the concept: form precedes force, which precedes matter, which precedes sensation, memory, communication, volition, conduct, and so on. From this we know that if a link referent is found only in low-numbered thesaurus entries, it will be more abstract than referents represented by high concept numbers.

Axis 2: Establishing the referent's applications from the *quantity* of concepts

The word "for" is referenced in 83 separate categories in the thesaurus. This is a large number of entries, and if you didn't know what the word meant, or how it could be used, you'd be no better off after having consulted the 83 references. Any term defined in so many different ways cannot be understood literally. On the other hand, there are 144 instances of the word "assignment" in the thesaurus, yet these all occur within only 4 conceptual categories. This suggests that the term is widely used, but its usage is not unclear. In fact, as an operational term, the word "assignment" is very clear, despite the variety of contexts in which it is employed. These examples suggest that terms that are overloaded with many connotations are more abstract, less literal. Therefore, the number of thesaurus concepts in which a link referent is found is a direct measure of how literal or abstract that referent might be. I have found it useful to draw the line at around 20 shared concepts, but such a threshold more directly depends on the content of the hypertext under investigation, since the measure of abstraction may be different in literary hypertexts than in scientific ones.

Axis 3: Establishing the referent's connotations from concept *range*

Beyond the number of concepts in which a link referent can be found is the range or span covered by those concepts. Whether such a referent resides in 4 thesaurus concepts or 40, there is also the question of the proximity of the concepts, their semantic neighborhood, one might say. To use a previous example, the four categories of term "assignment" are widely dispersed, being located in both the abstract and the concrete pole of the concept spectrum. The term has greater emphasis on its concrete connotations (thesaurus numbers 155: Assignment of cause, attribution; 755: Vicarious authority, commission; 783: Transfer; and 786: Apportionment). Taken with the fact of its 144 occurrences throughout the thesaurus, the spacing of the 4 categories into which it falls, indicating that it is widely used, but not categorical in itself. Therefore it likely points to some very specific lexia. On the other hand, link referents that possess a *narrow* range of concept numbers indicate terms that are precise or closely circumscribed. An example here is the word "fiction," which exists primarily in three categories (thesaurus numbers 515: Imagination; 546: Untruth; and 594: Description). This term's conceptual proximity is clear from the relative compactness of its semantic space: 515, 546, 594 designates 3 connotations with a span of 75 concepts out of the thesaurus's 1,000. This indicates that the referent has few connotations in descriptive usage, and is therefore useful as a link referent because it would point to some destination whose specificity to the overall hypertext topic can be assessed before entering such a node.

In these three basic metrics, we are really combining statistical variance with regression, aided by the mapping of a large space of categorical abstractions. The real value of a paratextual inference is the interplay of all three dimensions in the consideration of a hypertext's set of link referents. A paratext requires a small number of easily obtainable ingredients. One is a hypertext, the next is a broad semantic space, such as a Roget's Thesaurus. Certainly any such conceptual space can also be constructed for more special purposes, as Trigg's taxonomy indicates, and it is always best when the concept topography is laid out in a spectral sequence (e. g., abstract to concrete) as well as numbered. Lastly, any development environment will do. Every iteration of this experiment has been implemented in Perl, which accelerated designs, implementations and refinements.

WHAT AWAITS

Numerous directions are possible from this early point; all revolving around the ongoing exploration of what account links, if organized as a metropolis of meaning, could give of a hypertext, that other elements alone cannot. The key speculative step is in organizing and interrogating links as a critical mass of signification that I am calling the paratext. In special purpose contexts, such as in journalistic reporting or medical patient history and diagnosis, the inevitable pre-existence of information categories potentially makes the paratext a tangible partner in augmented interpretation. But literary hypertexts are probably the first and best context for paratextual analysis.

The paratext (unrelated to Genette's more recent use of the term for footnotes and other traditional book indicia) makes sense when the link referents are at least minimally descriptive, and when they are mapped out over a semantic space that can help interrelate them. The approach is in some incipience, so weaknesses are fairly clear: link referents built as *phrases* or metaphors will not be analyzed atomically or with their intended meaning, but as separate words. A paratext should also be used alongside other structural methods for assessing the use and quality of links. It doesn't obviate the commonsense questions and practices of structure and clarity of navigation that must inform good hypertext design.

The value of any approach is in its ability to resolve questions not answerable by others. In use, a paratext can determine the abstractness, connotation, and application of link referents, enabling the interrogation of links as a kind of narrative in hypertexts. Possible further work might involve determining the most common categories of a hypertext's link referents as a measure of readability, convergence, or clarity in the use of links within a hypertext. Large-scale organization of paratextual axes—abstraction, connotation, and linguistic application—as applied to larger corpora could also reveal unforeseen patterns of signification at the link referent level. Perhaps the paratext may help reveal patterns that remain hidden from either the perspective of linear text or closed hypertext systems.

Overall, in ways very different from the classical rhetoric and narrative, what emerges from a third epoch of writing, writing after the age of sequence, might produce its own touchstones for intertextual understanding. Here, the paratext might pose for us central questions of the text, Gadamer's key to hermeneutic understanding. And in this possibility, the paratext can invite interpretations that argue more than anything for the richness of this layered form that is the covenant of hypertext alone.

ACKNOWLEDGMENTS

REFERENCES

1. Aarseth, E. Cybertext: Perspectives on Ergodic Literature. University of Bergen, Bergen, 1995.
2. Allan, J. "Automatic Hypertext Link Typing." *Proceedings of Hypertext '96*, (March 16-20, 1996).
3. Andrews, K., Kappe, F and Maurer, H. Serving information to the Web with Hyper-G. *In*: Computer Networks and ISDN Systems, 27(6), pp. 919-926, 1995.
4. Barthes, R. The Semiotic Challenge. University of California, Berkeley, 1985.
5. Bernstein, M. "An Apprentice That Discovers Hypertext Links." *Proceedings of ECHT '90*, (November 27-30, 1990).
6. Brigham, L. Cyberinthian Ways. Electronic Book Review, (Winter, 95/96) http://altx.com/ebr/ebr1/brigham.htm
7. Charney, D. "Comprehending Non-linear Text: The Role of Discourse Cues and Reading Strategies," in Hypertext '87 Papers, November 13-15, 1987, University of North Carolina, Chapel Hill, 1987.
8. Clitherow, P., Riecken, D., Muller, M. "VISAR: A System for Inference and Navigation in Hypertext." *Hypertext '89 Proceedings*, (Nov. 5-8, 1989).
9. Conklin, J. Hypertext: An introduction and survey. Computer 20, 9, (Sept. 1987).
10. DeYoung, L. 1990. "Linking Considered Dangerous." *Proceedings of ECHT '90*, (Versailles, France, Nov. 27-30, 1990) pp. 238-249.
11. Gadamer, H.-G. Truth and Method. Continuum, New York, 1961.
12. Hayes, P, Pepper, J. "Towards an Integrated Maintenance Advisor." *Hypertext '89 Proceedings*, (Nov. 5-8, 1989).
13. Joyce, M. Of Two Minds: Hypertext Pedagogy and Poetics. University of Michigan Press, Ann Arbor, 1995, p. 70.
14. Kristeva, J. Semiotikè: Recherches pour une sémanalyse. Seuil, Paris, 1969.
15. Landow, G. Hypertext 2.0: The convergence of contemporary critical theory and technology. Johns Hopkins, Baltimore, MD, 1997, p. 97.
16. Levy, D. M., Marshall, C. C. "Going Digital: A Look at the Assumptions Underlying Digital Libraries." *Communications of the ACM*, 38, 4, (1995), pp. 77-84.
17. Marshall, C. C., Shipman, F. M. "Searching For The Missing Link: Discovering Implicit Structure In Spatial Hypertext." *Proceedings of Hypertext '93*, (November 1993).
18. Nelson, T. Literary Machines. Mindful, Swarthmore, 1990.
19. Schmandt-Besserat, D. The earliest precursor of writing. *In*: Language, writing and the computer. W. S-Y Wang, Editor. W. H. Freeman & Co., New York, 1978.

20. Smith, K. E., Zdonik, S. B. "Intermedia: A Case Study of the Differences Between Relational and Object-Oriented Database Systems". *OOPSLA '87 Proceedings*, (1987), pp. 452-465.

21. Trigg, R. A Network-Based Approach to Text Handling for the Online Scientific Community. Ph.D. dissertation. University of Maryland Technical Report, TR-1346. University Microfilms #8429934, College Park MD, 1983.