

## Hypertext and Deixis

Dan Loehr  
Georgetown University  
loehrd@gusun.georgetown.edu

### Abstract

Hypertext linking is akin to deixis in that each points out of the discourse to the context in which the discourse occurs. The context of a hypertext document such as a World Wide Web page is the World Wide Web itself. This study investigates some linguistic aspects of hypertext, using as data Web pages from college students and commercial enterprises. I first look at the syntactic forms of hypertext anchors, finding them mostly noun phrases. Next, noting the intuitive similarities between hypertext and place deictics, I attempt to find a quantitative relationship between the two. I show that certain place deictics do occur more frequently inside of hypertext anchors than in the ordinary text of Web pages, but this is due to the popularity of the deictic *here*, as in *click here*. Observing that hypertext style guides prescribe against the usage of *click here*, I hypothesize that more experienced hypertext authors will disprefer this usage. Using the assumption that college students are less experienced than commercial Web authors, I am unable to confirm this hypothesis statistically. Finally, I examine the various types of deixis Web authors use in hypertext anchors, finding the majority to be person deictics. I close by noting that the variety and creativity of deixis within hypertext anchors suggests that the relatively new medium of hypertext is a fully productive genre of natural language.

### 1. Introduction

Deixis is, in Lyons' [1979] simple phrase, "identification by pointing". Specifically, deictic terms point out of the current discourse to facts concerning the context of the utterance. Deictics include words such as *I*, *here*, and *now*. Sentences

with these terms can not be understood fully unless one knows the people, place, and time involved in the utterance of the sentence. As an example, imagine finding a bottle in the ocean with the note, "Meet me here at noon tomorrow" (from Fillmore [1975]). Unless we know when, where, and by whom the note was written, we are at a loss to interpret the sentence.

With the evolution of hypertext, we find another type of deixis. Hypertext, such as is found in World Wide Web pages, is electronic text which "points" to other locations in the same or different electronic documents. When hypertext is selected by means of a mouse or keyboard, the document "pointed to" is retrieved and displayed on the screen. An example of hypertext might be: *I am a student at Georgetown University*. (The hypertext is underlined, a convention I will follow in this paper). If a reader of that sentence clicked on the words *Georgetown University*, an electronic document pointed to by the hypertext would be displayed: presumably (but not necessarily) relating to Georgetown University.

This, then, is also deictic. Just as ordinary deictics can not be fully evaluated without leaving the utterance and looking to the outside world, neither can hypertext be fully evaluated without leaving the text and looking at its context (the World Wide Web itself) to see what is being pointed at. We might call this hypertext deixis *hyperdeixis* (term credited to Dr. Cathy Ball), referring to the fact that hypertext "points", just as natural language expressions such as *I*, *here*, and *now* point. It is interesting to note that hypertext links are often referred to as "pointers", while the word *deixis* itself is from the Greek word for *pointing*.

It is important to make a distinction here. The deictic aspects of hypertext need not have *anything* to do with the words used as the *anchor* of the hypertext (i.e., the words which, when selected, activate the link). The deictic aspect has solely to do with the fact that the hypertext link points elsewhere. The words in the anchor do not have to relate to where the link takes you - in fact, they do not have to be words at all. The anchor can be a picture, or simply a blank spot on a page. No linguistic analysis of the anchor is needed to follow the hypertext link.

Little is known about the linguistic properties of hypertext. Although many questions present them-

selves, this paper will look at four:

- (1) Given that hypertext - text which "jumps somewhere" - is a relatively new phenomenon, just what forms do authors use for hypertext anchors?
- (2) Given the intuitive "pointing" similarity between *place* deictics (expressions which point to a place) and hypertext, is there any relationship between the two? Specifically, do authors of the new medium of hypertext tend to use place deictics more frequently inside of hypertext anchors, since hypertext "reminds" them of place deictics? Or, for the same reason, do authors tend to *avoid* using place deictics inside of hypertext anchors?
- (3) If any relationship is found in (2) above, do writers' preferences for using place deictics in hypertext *change* as they become more experienced with this new genre?
- (4) Again noting the similarity between place deictics and hypertext, are place deictics the most common form of deixis found within hypertext anchors? Or are other types, such as person deictics or time deictics, more prevalent?

This paper, then, represents an exploratory foray into the little-known world of hypertext and deixis. It will undoubtedly generate more questions than answers.

## 2. Background

The earliest major work on deictics (also called indexicals) is by Bühler. He begins with an elegant image:

The arm and the finger gesture of man, to which the index finger owes its name, recurs when the signpost imitates the outstretched "arm"...and the question [is] posed as to whether spoken language contains signs that function as signposts. The answer is yes, deictic words such as *here* and *there* have a similar function. [1934]

Bühler did not originate the idea of deixis. He notes

that classical Greek grammarians were aware of the concept. In the modern era, he credits turn-of-the-century linguists such as Brugmann and Wegener, who, while reconstructing Proto-Indo-European, described four types of deixis: *I*, *you*, *this*, and *that*. Bühler notes that *I* and *you* are referred to as *personalia* (Latin *persona*), which is from the Greek for *mask* or *role*. The meaning is that *I* and *you*, like actors, switch roles. Bühler then advances a two-field theory: a *deictic field*, in which deictics point to the actual entity in the world, and a *symbolic field*, in which all other terms refer symbolically to entities in the world.

Fillmore [1975] gives a very thorough taxonomy of deixis. He enumerates five types: person, place, time, discourse, and social.

*Person* deixis refers to the speaker, addressee, and audience of the utterance. Examples are pronouns and possessives such as *I*, *we*, *yours*, *my*, etc.

*Place* deixis, including the locative adverbs *here* and *there* and the demonstratives *this* and *that*, are divided into two types. *Gestural* place deixis requires the addressee to physically monitor the utterance for an accompanying gesture: "Put the flower pot *here*". *Symbolic* place deixis requires only knowledge of the conditions of the utterance, without physical monitoring. My friend from London may write, "It's raining *here* again." Though I did not see her write the letter, I know where she was when she wrote it, and thus I know where *here* is. As an interesting sidelight, Fillmore notes that there is only one word in English which *requires* gestural deixis for its use: *yea*, as in "She is *yea* tall".

*Time* deixis, including the adverbs *now*, *last night*, and *tomorrow*, involves three different times. The *encoding* time is the time at which the utterance is made. The *decoding* time is the time at which the utterance is received (which will be delayed in written text). Finally, the *reference* time is any time referred to in the topic of the sentence. A typical voice-mail message might illustrate all three types. "This is Dan calling at 9 AM (*encoding time*). Give me a call back when you get this message (*decoding time*). I'd like to set up a meeting for Tuesday the 14th (*reference time*)."

*Discourse* deixis, according to Fillmore, "has to do with the choice of lexical or grammatical elements which ... refer to some portion ... of the ongoing discourse -- something like, for example, 'the former'." Discourse deixis (also called *textual deixis* by Lyons [1979]) can be both written ("This sentence is written in English") and spoken ("She spoke about *this* loud"). Fillmore notes that discourse deixis is de-

scribed temporally, as "any point in discourse can be thought of as a point in time". For this reason, we can say "in the *preceding* paragraph", or "the *following* announcement is a paid advertisement". In addition, written discourse has unique terms due to pages being read top-to-bottom: thus, we say, "as mentioned *above*", or "see *below*".

Finally, *social* deixis encodes, in Fillmore's words, "the social relationships on the part of the participants in the conversation that determine, for example, the choice of honorific or polite or intimate or insulting speech levels". In addition to honorifics, examples include nicknames, informal names, and titles of respect.

Fillmore also notes that certain verbs, such as *come* and *go*, are also deictic, as they imply a *here* and a *there*.

Considerable work in the field has also been by Lyons. He makes note of the *deictic center* - the "zero-point of deictic context", which is always egocentric for the speaker [1968, 1977]. This zero-point is in relation to person, time, and space, with the center representing *I*, *here*, and *now*. Deictic terms are expressed in relation to this deictic center. *Here* and *this* are closer than *there* and *that*, and *now* is closer than *tomorrow*. It is illustrative that the *first*, *second*, and *third* person radiate outward from the deictic center. Lyons reminds us that deictic terms may also have non-deictic uses [1973]. "*This* idea is not new" refers anaphorically to some previously expressed idea, while "*This* book is dedicated to my husband" refers deictically to the book containing the dedication. Finally, Lyons discusses tensed verbs as time deictics [1977]. Note that we need to know the time of utterance to work out any truth conditions of a tensed sentence.

I have thus far summarized work on deixis. As for deixis in relation to hypertext, no previous work exists that I am aware of. I hope that this study (a deictic!) will shed some light on this new phenomenon.

### 3. Forms of Hypertext Anchors

Before addressing my first question - the forms authors use for hypertext anchors - I will briefly discuss my general methodology. Methodology specific to each of the other three questions will be covered in that question's section, in turn.

The data used for all four questions consists of pages

from the World Wide Web collected in April 1996. There are of course many types of Web pages (Roberts [1997] discusses the different genres found). For this study, I collected pages from two types: the "home pages" of college students, and the "home pages" of commercial enterprises. A home page is the initial, top-level page at a "web site", with a web site typically consisting of multiple related pages. For consistency, I used home pages only.

For college student data, 126 different Web pages (containing roughly 1000 anchors) were collected randomly from three institutions: Cornell University, the University of Texas at Austin, and the University of Oregon. For commercial data, 31 Web pages (containing roughly 500 anchors) were collected. The commercial businesses included well-known firms in varying fields: examples are Apple Computer, Prudential Insurance, Pepsi-Cola, Sony, Whirlpool, Disney, and the American Stock Exchange.

After each page was downloaded, an automatic script stripped out the words used as anchors. (Fortunately, Web pages are written in HyperText Markup Language (HTML), which has specific tags surrounding the anchors). The anchors were saved to a separate file, as was the original Web page, for reference. Graphical anchors were ignored, unless they had "alternate" text associated with them, for the use of non-graphical Web browsers.

To answer question (1), I then went through the list of anchors, and manually coded them per the following categories:

- Noun Phrase
- Verb Phrase
- Adjective Phrase
- Adverb Phrase
- Prepositional Phrase
- Relative Clause
- Full Sentence
  - Imperative
  - Question
  - Declarative
- Other

The results are shown in Table 1, which lists in roughly decreasing order the forms authors used for hypertext anchors, by number and percentage. The table is further broken down by source (college student and commercial), for comparison. Note that the "Full Sentence" category has been subdivided into "Imperative", "Question", and "Declarative".

Form	College Student Home Pages		Commercial Home Pages		Examples
	#	≈%	#	≈%	
NP	893	88.9	364	72.4	<a href="#">Cornell University Home Page</a>
Full Sentence	48	4.8	69	13.7	
Imperative	32	3.2	41	8.1	<a href="#">Sign-in to Win!</a>
Question	9	0.9	15	3.0	<a href="#">What's New?</a>
Declarative	7	0.7	13	2.6	<a href="#">General Motors Announces Its Electric Vehicle</a>
VP	19	1.9	26	5.2	<a href="#">click here</a>
AdvP	18	1.8	4	0.8	<a href="#">click here</a>
AdjP	10	1.0	7	1.4	<a href="#">New at National</a>
PP	8	0.8	12	2.4	<a href="#">About Me</a>
Other	7	0.7	4	0.3	<a href="#">If Operating Systems Were Beers...</a>
Relative Clause	2	0.2	17	3.4	<a href="#">Who We Are</a>
Totals	1005	≈100	503	≈100	

Table 1. Forms of hypertext anchors, number and approximate percentage, by source (Percentages do not exactly total 100.0, due to rounding)

We can now answer our first question, as to the forms hypertext anchors take. The overwhelming majority are noun phrases, with full sentences taking a distant second. Note however, that commercial Web authors tend to use fewer noun phrases and more full sentences than college students (the difference is significant, per a chi-squared test, with  $p \leq 0.01$ ). Perhaps this is because commercial sites try to present a more polished image.

Our first answer regarding hypertext immediately raises another question: Why are the overwhelming majority of anchors noun phrases? An intuitive reply is that authors are viewing hypertext anchors as *referring expressions*. Since referring expressions in more familiar *non-hypertext* are almost always noun phrases, authors choose noun phrases for their anchors. Continuing this theory, to the extent that authors are *not* using noun phrases as anchors, to just this extent are they "branching out" in the new phenomenon of hypertext, and discovering linguistic forms other than noun phrases for encoding hypertext links. Confirmation of these speculations, however, must await further research.

#### 4. Relationship Between Place Deictics and Hypertext

Let us now turn to our second question, reprinted here:

- (2) Given the intuitive "pointing" similarity between *place* deictics and hypertext, is there any relationship between the two? Specifically, do authors of the new medium of hypertext tend to use place deictics more frequently inside of hypertext anchors, since hypertext "reminds" them of place deictics? Or, for the same reason, do such authors tend to *avoid* using place deictics inside of hypertext anchors?

To determine this, we need to determine the frequency, or "density", of place deictics *inside* of hypertext anchors versus that *outside* of hypertext anchors (i.e. in the regular text of the Web page).

A digression is needed to motivate my methodology for determining these frequencies. At first blush, the task seems relatively easy. One need simply go through the Web pages and code for each occurrence of a place deictic. Although time-consuming, this can be aided by heuristics such as searching for the most common place deictics: *here*, *there*, *this*, and *that*. Next, a simple programming script can split apart the Web pages into two files: one file containing all the words inside of hypertext anchors, and the other containing all the words outside of the hypertext anchors. Finally, one can calculate the relative

frequency of the place deictics in each file, by dividing the number of place deictics by the number of words in the file.

There is a problem, however, in determining the frequency of something by dividing by the number of words in a text. As Ball explains:

In a word-based frequency analysis, to say that a phenomenon occurs with equal relative frequency in two samples is to say that equal amounts of text, measured in words, will yield the same number of tokens. But relative frequency should be a measure of the number of times something occurs within the number of *opportunities for it to occur* (emphasis mine) [1994].

To give an example: If we are comparing the frequency of, say, interrogative sentences in the writing styles of Hemingway and Dostoevski, it would be meaningless to divide the number of interrogatives by the number of words in each author's text. The results of this would be skewed by the *length of sentences* which each writer produces. Hemingway may pen fully twice the number of sentences in the same number of words as his Russian counterpart. Thus, if both authors had the same absolute number of interrogatives, we would be wrong to conclude that they both favored this form equally. Rather, Hemingway, who produced twice as many sentences, used the interrogative only half as frequently as Dostoevski. The key, then, is to measure a phenomenon against the number of opportunities for the phenomenon to occur, not against raw word count. In this example, the opportunity for an interrogative to occur is a sentence - hence, we must count the number of sentences as well.

We may now return to our problem of measuring the frequency of place deictics. We must decide what an appropriate *opportunity for occurrence* for a place deictic would be. As Ball advises, selecting this category is "crucial and requires considerable thought [1994]", or else our results will be invalid.

With this in mind, let us look first at the place deictics *here* and *there*. Although technically these may be thought of as adverbs ("Come *here*"), I will argue

that they *pattern* with prepositional phrases ("Come to *this place*"). As further evidence, consider "Put the book *here*". Replacing *here* with an adverb is ungrammatical (\* "Put the book *slowly*"), while a prepositional phrase is acceptable ("Put the book *on the table*"). Thus, when an author chooses to use *here* or *there*, the author is choosing to use these words *instead* of using a prepositional phrase. Hence, I will consider an opportunity for occurrence of *here* and *there* to be a prepositional phrase.

Although searching for prepositional phrases is time-consuming, it is certainly doable - and in fact, I did just that for this study. As a heuristic, I started by searching for a list of 50 common prepositions, which I estimate covered at least 90% of all raw occurrences of prepositional phrases (the first six alone accounting for 70%, according to Mindt and Weber [1989], reported in Crystal [1995]). I then visually scanned the text for any remaining prepositions. (A more efficient method would be to scan for the 81 prepositions found in both the Brown and LOB corpora, said by Mindt and Weber to account for 99.9% of all raw occurrences).

We must also consider the other types of place deictics. While sometimes acting as determiners, the place deictics *this* and *that* can also be (and pattern with) both pronouns ("Look at *that*") and adjectives ("*This* morning"). Unfortunately, this means that an opportunity for an author to use *this* or *that* would be a noun phrase or an adjective phrase. While searching for prepositions is difficult, it is feasible because prepositions are a finite class. Yet searching for the open classes of nouns and adjectives I judged to be undoubtably time-consuming for this research. Equally difficult would be considering place deictics occurring as motion verbs (i.e. *come*), as that requires searching for *all* verbs to find opportunities for occurrence. I have therefore chosen to restrict my study of place deictics to only *here* and *there*. Hence, my research will have nothing to say about other place deictics. A strategy for resolving this dilemma would be to use an automatic parser, which could identify most nouns, adjective, and verbs. I leave that task to future work.

Having covered the necessary background for comparing the frequency of place deictics inside of and outside of hypertext anchors, I may now present the results in Table 2:

	Inside of Hypertext Anchors	Outside of Hypertext Anchors	Totals
Number of Occurrences of <i>Here</i> and <i>There</i>	42	66	108
Number of PPs (Opportunities for Occurrence of <i>Here</i> and <i>There</i> )	577	2433	3010
Frequency (in Percentage)	7.3	2.7	
Totals	619	2499	3118

Table 2. Frequency of the place deictics *here* and *there* inside of and outside of hypertext anchors

In percentages, the place deictics *here* and *there* occur at a frequency of 7.3% inside of hypertext anchors, versus 2.7% outside hypertext anchors (the difference is significant, per chi-squared with  $p \leq 0.001$ ). For simplicity, I have grouped the educational and commercial pages together in this table - broken apart, there is no statistically significant difference between them in this regard.

Thus, based on the place deictics I've looked at, we can answer our second question: place deictics do occur more frequently inside of hypertext anchors than in the regular Web page text.

However, the conclusion that this is so because of the place-deictic quality of hypertext is not so easily reached. It turns out that of the 42 place deictics found in anchors, 41 of them were the word *here*. This abundance is due to the popularity of the hypertext expression *click here*. This usage may be a simple Graphical User Interface (GUI) issue, in which authors are merely instructing their audience to click a certain button. Furthermore, if authors want their readers to go to some other place *there*, it is counterintuitive to place deictics to use the word *here*. Thus, the usage of *here* in hypertext is probably not due to place deictics, but rather a GUI artifact.

##### 5. Changes in Hypertext Authors' Style as a Function of Experience

In answering question (2), we could not prove a relationship between hypertext anchors and place deictics. We did, however, see a relationship between the word *here* and hypertext anchors. Going with this, let us modify our third question to be:

- (3) Do writers' preferences for using the word *here* in hypertext *change* as a function of experience in this new genre?

This question is not unfounded. Many hypertext style guides preach against what has been labeled the "click-here syndrome". The argument is that the word *here* has nothing to do with the target of the link, and should be avoided in favor of a more descriptive term. To quote one HTML manual:

**STYLE TIP:** Try to create links that don't muck up the flow of the text. ...Your text should stand on its own. You might want to avoid words like "click your mouse on this sentence to read about the history of skyscrapers." Every Web user knows how to click. ... Instead, say something like, "There's an excellent article on the history of skyscrapers by T.O. Tall" - and make the word "skyscraper" your link. (Fox and Downing [1995])

If this is what is prescribed, it is reasonable to hypothesize that professional writers will try to avoid the click-here syndrome. To test this, we can look at the frequency of the word *here* within the hypertext anchors of novice writers versus that of professional writers. To this end, I make the following assumption - that college students are less experienced at writing Web pages than those who author commercial pages. This is of course a generalization, but given the fact that commercial outfits are concerned about the image they present, and that their Web page authors are most probably paid for their work (and hence professionals), it seems reasonable that they will hire experienced authors. No experience, however, is required for college students' authorship.

Table 3 presents the frequency of *here* in the hypertext anchors of college student home pages versus commercial home pages.

	College Student Home Pages	Commercial Home Pages	Totals
Number of Occurrences of <i>Here</i>	34	7	41
Number of PPs (Opportunities for Occurrence of <i>Here</i> )	512	64	576
Frequency (in Percentage)	6.6	10.9	
Totals	546	71	617

Table 3. Frequency of the term *here* inside of hypertext anchors for college student home pages versus commercial home pages

At first glance, the opposite of our hypothesis seems to be true - that the more experienced commercial authors use *here* more frequently. However, the difference is not significant (per chi-squared at  $p \leq 0.05$ ), so any discussion is moot. Hence, I cannot confirm or reject the hypothesis based on this data. Greater quantities of data may be necessary to determine any real patterns. In addition, future work might also employ a more verifiable method of identifying novice and experienced Web authors than simply generalizing about college students and commercial webmasters.

## 6. Types of Deictics Found in Hypertext

Finally, let us turn to our fourth question:

- (4) Again noting the similarity between place deictics and hypertext, are place deictics the most common form of deixis found within hypertext anchors? Or are other types, such as person deictics or time deictics, more prevalent?

To answer this, I looked at each anchor in the data to see if it contained any deictics. These deictics were then categorized as one of:

- Person
  - Speaker
  - Audience
  - Both
- Place
- Time
- Discourse
- Social
- More than one of the above

Some discussion is necessary regarding my coding. First of all, the common *Click here*, as noted, was classified as a place deictic, as it refers to a physical

location in the universe (a spot on a computer screen which may be clicked on). Yet as a sidelight, one could also consider it a discourse deictic, as it refers to the word *here* itself, which is part of the discourse. This interesting case would be the purest discourse deixis imaginable. Most discourse deixis refers to the text *in which* the deictic word occurs (*this* book, the paragraph *below*). We even have cases where sentences refer to themselves (*This sentence is in English*). But the deictic *here* in *click here* would be a word which points to itself. Verbally, it is difficult to find an equivalent. One possibility would be, "The word I say *now* has three letters and rhymes with cow". Though certainly possible before hypertext was developed, it is the advent of hypertext which has made this "ultimate discourse deixis" commonplace in language.

(Note, however, that person deixis has always had this reflective deixis, at least verbally, in the first person. The speaker who says "I" is referring to herself. Thus, the word *here* (in a manner somehow reminiscent of Lewis Carroll) is *personified*, by calling attention to *itself*..)

There were cases where *here* was not used in the *click here* sense. One example was an anchor labeled *The weather here*, referring to the weather at Cornell University, the author's home. Another was the following: *Here I am in Roslyn, WA, (where Northern Exposure was filmed)*. The hyperlink brought up a picture of the author in Roslyn, WA. This *here* is not referring to the word *here*. Nor is it referring to the location in which the author wrote the sentence (at Cornell University, in Ithaca, NY). Rather, we have a case in which the author has shifted his deictic center mentally, much as Thomas [1996] reports that speakers do when giving directions.

In terms of person deixis, I collapsed the *addressee*

and *audience* types under the latter category. Though typically *audience* refers to others who hear the utterance (in addition to the intended addressee), Web pages are intended for everyone to read, not a specific addressee. Thus, there are no "bystanders" - or, more accurately, the bystanders *are* the addressees. Thus, for the Web, addressee and audience are the same thing. I have chosen the latter category as more accurately reflecting how the Web works.

Lyons' notion of deictic and non-deictic uses of words was needed in several cases. For instance, *this* in *Why is this page black?* is clearly deictic. But in another anchor, *Come to Think of It. Who Is This Phil Guy?*, the same word is not. Rather, it is anaphoric, since Phil has introduced himself at the beginning of the Web page.

Finally, when coding, I resisted the temptation to collect implied deictics. These occurred in two forms. The first was the imperative: an example was

*Send an Electronic Postcard*. Imperatives imply the second person, as in *(You) Send an Electronic Postcard*. The second form of implied deictic was the implied first person possessive, as in *Favorite Web Sites*. It is probable that the author meant *My Favorite Web Sites*. However, though both of these types implied a person deictic, the actual deictic lexical item was missing. There were other cases where a deictic word (*my*) was just outside the anchor, as in *My father*. In all these instances, since there was no deictic word in the anchor, I did not classify it as containing an ordinary deictic.

Now, to answer our fourth question. Table 4 lists the distribution of deictics (by number and percentage) found in hypertext anchors.

Type	Number	Approx. Percent -age	Examples
Person Only	69	54.8	
Speaker	60	47.6	<u>Mail Me</u>
Audience	4	3.1	<u>your feedback</u>
Both	5	4.0	<u>Tell Us Your Story</u>
Place Only	38	30.2	<u>Return to Jane Austen's info page</u>
Discourse Only	10	8.0	<u>Why is this page black?</u>
Multiple Types	7	5.6	<u>alt=Make your visit count. load this image</u>
Time Only	2	1.6	<u>I'm in the lab</u>
Social Only	0	0.0	
Totals	126	≈100	

Table 4. Types of deictics found in hypertext anchors (Percentages do not exactly total 100.0, due to rounding)

For simplicity, I have again grouped the educational and commercial pages together in this table - broken apart, there is no statistically significant difference between them on this question.

We can now answer our fourth question. Of those deictics used in hypertext anchors, the majority are not place deictics, but person deictics. And of these, the majority refer to the speaker. Of course, the data for this effort was home pages, and the preponderance of self-references may reflect the fact that home page authors use their home page as a vehicle to tell the world about themselves. As research into the various genres of Web pages progresses, it will

be interesting to see which is the most common type of home page. This data suggests that personal narratives may be the most common.

Though we have answered our fourth question quantitatively, there remain some interesting qualitative examples to discuss.

First of all, following Fillmore's sense of *come* and *go* as deictic verbs, the common anchor *return* is also deictic, as in *Return to Jane Austen's info page*. To *return* implies a starting point - a *here* - to return from. Though the place we are "going" is specified (Jane Austen's info page), the place we are



"leaving" (here) is not. We can only know where we are returning from by deixis - by knowing the place of utterance. Note that this is another case where the deictic center is mentally shifted. The reader never actually moves anywhere - she is still in her chair before her computer. But mentally, she is traveling to wherever her Web documents are retrieved from. Such is the magic of the World Wide Web.

Another interesting area has to do with time deictics. There were only three of these found in anchors (two on their own, and one in combination with other deictics). The first rather novel example had a blinking anchor, in the sentence "Click now!" Perhaps the author meant for the reader to click at the moment the blinking anchor was lit up. In the second example, the author writes, *I may be in the computer lab, in fact, you can check if I'm in the lab right now.* When this anchor is selected, a snapshot comes up which is taken from a live video camera in the author's laboratory (the author was not in when this data was collected). Even though the word *now* is not within this anchor, we can still find time deixis in the tensed verb *am*. The present tense implies just that - *now* - and the sentence cannot be evaluated unless one knows the time it is read. Note that both of these examples use decoding time (the time the sentence is read).

The third, very clever, example is:

I'd like to send a special hello to Stephen Crawford. Why don't you, too? And watch here for a link to his future home page. Then you, too, can meet the greatest trombonist in the world

Still watching?

His home page isn't future anymore, it's here!

With his final *here*, this author has scored a deictic triple play. First, he invokes *here* as a place deixis, meaning *Stephen Crawford's home page is now available from this place here*. (Note the sentence above - *Watch here for a link* - this can be taken to mean *Watch this place on the Web*). Second, he uses the familiar *click-here* phrase - which, as mentioned, could be thought of as discourse deixis. Finally, the author uses a sense of *here* which means *now*, as in *It's December 25, and Christmas is here*. This invokes the third deixis (time). Note that this example uses encoding time (the time the sentence was written).

Another interesting point regards gestural place deixis. One may think that gestures (actual pointing) are not possible in a textual medium. Yet Web pages are more than text - they support animation, and many modern pages contain "moving arrows" acting as anchors. Even non-graphical, text-based pages can create arrows such as these: --> <--. However, I concentrated only on textual anchors (avoiding the difficult problem of classifying images), and found no gestural deixis in my data. Yet temptingly close was the sentence *Here I am* (which brought up a picture of the author). This might be thought of as a presentative, much like the French word *voilà*, which requires a gesture for its use. The words *Here I am*, by convention, almost seem to imply a gesture - the one of an actor stepping from behind a curtain, sweeping his hands before him and bowing his head in presentation to the audience.

Note that no occurrences of social deixis were found.

## 7. Conclusions

Hypertext links behave much like ordinary deictics. Both require the context of the utterance for full interpretation. For hypertext, the "utterance" is the Web page, and the context of this Web page is the World Wide Web itself. Just as ordinary deictics like *me* make us look to the context to see who *me* really represents, so does hypertext make us look to the context of the Web to see what the hypertext link really represents.

This study investigated several questions regarding hypertext and deixis, using as data Web pages collected from the Internet. In terms of form, hypertext anchors tend to be noun phrases, with a lesser concentration of complete sentences. Although one might intuit a relationship between place deictics and hypertext, none could be substantiated. Hypertext anchors do contain a greater concentration of a certain place deictic (*here*) than the rest of the Web page, but this is due to the popularity of the phrase *click here*. This popularity, in turn, may be due to the unique environment of hypertext, which makes the text itself a computer input "button".

Because style guides prescribe the avoidance of the "click-here syndrome", it is reasonable to hypothesize that more experienced hypertext authors will disprefer this in contrast with less experienced authors. Assuming that college students are less experienced than professionals who author commercial Web pages, I was unable to confirm this statistically. As to the types of deixis found inside of hypertext anchors, the majority were person deictics referring to the speaker, probably due to the fact that the pages studied were home pages, in which authors

present themselves to their audience.

The only statistically valid difference between the two sets of authors is that professional authors tended to use more full sentences and less noun phrases than college students. This may have nothing to do with deixis. It may instead be due to the polished image which commercial sites hope to project to their customers.

Further work in this area might include the following. Our first question revealed that the majority of hypertext anchors were noun phrases. This might be investigated with an eye to the use of anchors as referring expressions, which are typically noun phrases. The second question dealt with determining the relative frequency of a variety of place deictics, noting the difficulty of calculating their opportunities for occurrence (which manifest as a variety of parts of speech). A robust parser might be used to identify these parts of speech. The third question - whether experienced authors disprefer the "click-here syndrome" - remained unanswered. More data, and a more verifiable means of identifying experienced and novice authors, might prove fruitful. The final discovery - that the preferred deictics used in home page anchors were person deictics - might shed light on the classification of home page genres. My findings suggest that home pages are of a narrative form, utilizing first person.

Aside from any relationships between hypertext and deixis, we have seen that hypertext provides a unique environment for creative uses of deixis. Gestural deixis, time deixis, and even three-deixes-in-one-word were found. These examples, coupled with the fact that nearly all types of linguistic deixis are found in hypertext, lend evidence that the relatively new medium of hypertext is a fully productive genre of natural language.

### Acknowledgements

I gratefully acknowledge Hughes Electronics Corporation, for providing a fellowship under which this research was carried out. I am also indebted to Dr. Catherine Ball for the original idea of investigating hypertext and deixis.

### References

- Ball, Catherine. *Automated Text Analysis: Cautionary Tales*. Literary and Linguistic Computing, Vol. 9, No. 4, 1994.
- Bühler, Karl. *Theory of Language (Sprachtheorie)*. Translation of 1934 original by D. Goodwin. John Benjamin Publishing, 1990.

Fillmore, Charles. *Santa Cruz Lectures on Deixis*. Indiana University Linguistics Club, 1975.

Fox, D. and Downing, T. *HTML Web Publisher's Construction Kit*. Waite Group Press, 1995.

Lyons, John. *Introduction to Theoretical Linguistics*. Cambridge University Press, 1968.

Lyons, John. *Deixis as the Source of Reference*. Linguistic Agency University Trier, 1973.

Lyons, John. *Semantics, Vol. 2*. Cambridge University Press, 1977.

Lyons, John. Deixis and Anaphora. In T. Myers (ed.) *The Development of Conversation and Discourse*. Edinburgh University Press, 1979.

Mindt, D and Weber, C. *Prepositions in American and British English*. World Englishes 8, 1989. Quoted in Crystal, D. *The Cambridge Encyclopedia of the English Language*. Cambridge University Press, 1995.

Roberts, Greg. *Manuscript*. 1997

Thomas, Katie. *Manuscript*. 1996