Making Sense in the Margins: A field study of annotation*

James Blustein[1,2], David Rowe[2,3], and Ann-Barbara Graff[2,4]

[1] Dalhousie U. (Faculty of Computer Sci. & School of Info. Mgmt.) jamie@ACM.org
[2] Hypertext Augmenting Intelligent Knowledge Use (HAIKU) Project
[3] Dalhousie University (Faculty of Computer Science)
[4] Nipissing University (English Studies)

Abstract. We report on three years of data collected in the field from students in graduate and undergraduate seminars at two universities. The students annotated texts for discussion in classes where hypertext and computer interfaces were core topics. The results of our analysis show how annotation style changes with a combination of experience and study of material related to annotation. Our major conclusions are that there are essentially six purposes for scholarly user-readers to annotate; and support for textual glosses is a necessary part of any successful annotation technology for such use. Our study suggests tools that will be appreciated by e-text users.

1 Introduction

In academia, annotation is a means of making sense of complex material, marking engagement, navigating, and establishing a foothold for original thought. Wolfe and Neuwirth [27, p. 338] note that 'empirical research involving students suggests that annotations improve comprehension, facilitate rereading and reviewing of documents, and help writers bridge reading and writing practices'. To date, computer-based annotation tools lack something, insofar as they have been found to be less effective than their paper analogues for some purposes [15, 20]. Athough screen technologies are becoming more like paper other limitations, in particular the user interfaces of annotation software, impair wide-spread use [8].

We present an outline of the techniques that scholarly reader-users currently employ. Although annotation has been studied historically (e.g. by Jackson [12] and Hauptman [10]), practices have changed in recent generations. Mangen [15, p. 404] cites recent articles suggesting that 'reading modes and habits in general are changing due to steadily increasing exposure to digital texts'. Annotation however does not seem to have been affected by such exposure [20, 25].

* This article has been accepted for presentation at the *International Conference on Theory and Practice of Digital Libraries* (TPDL) in Berlin during September 2011. The original publication is available at www.springerlink.com.

2 Method

Rather than relying on prescriptive or historically derived taxonomies to identify engagement we follow Marshall and Brush's [20] lead in analyzing actual annotations created for known purposes and with recognizable motives.

We gave graduate and undergraduate students printed readings on topics related to hypertext with the instruction to prepare to discuss them in seminar in the following week. The only difference between these readings and regular ones was that these were presented in specially prepared form. Students were encouraged, but not required, to make annotations on the documents.

Materials All of the readings were presented on large pages[5] with wide margins, and ancillary material (e.g. diagrams and definitions) beside the text proper.

One reading was a collection of documents arranged around two related selections from a popular monograph about isochrestic design [22, 23]. The other reading was about the use of hypertext in education [21].

Participants and Activities The graduate students were enrolled in CompSci, Library and Information Studies, or Electronic Commerce at Dalhousie Univ. For many, English was not their first language. The undergraduates were enrolled in Literary Studies at Nipissing Univ. Of the 55 students enrolled between 2006 and 2009 fifteen volunteered to participate in the study ($N = 15$).

Students were graded on their discussion of the readings. Grad students were also graded on written critical assessments of the readings. Participation was not anonymous but the prof could not learn who had done so until after the course.

The undergrads studied Lipking's essay on marginal gloss [14] and Coleridge's annotated *Rime of the Ancient Mariner* [5] before the exercise. The only document they read for this study was the article about hypertext in education [21].

The grad students studied articles about annotation over the next weeks before a second document was circulated. They received grades and comments for their summaries of the first document before they could annotate the second.

Coding Method The annotations were categorized according to a taxonomy based on studies by Marshall et al. [17, 18, 20]. Each individual annotation was examined and its features (see below) were recorded. For annotations that were hybrid types each component was counted separately, i.e., if an annotation was composed of an arrow and a box then 3 annotations were recorded: compound, arrow, and box. The same procedure was used when tallying annotation function.

Annotation Category Annotations were coded as either telegraphic or explicit. A *telegraphic annotation* is a non-text based marking. Underlining, highlighting, and asterisks are examples of telegraphic annotations. An *explicit annotation* is a textual note (from single words to paragraphs).

[5] Ledger-size paper is 11"×17" — double the width of the letter-size (8½"×11") paper with which the students in North America were used to working.

Audience All annotations in our study were *private*, i.e. not intended to be read by others. Annotations for a *public* audience tend to be much lengthier and less spontaneous — they are not about working towards understanding but a performance to demonstrate mastery to later readers [19].

Location The place on the document where the annotation was recorded is its location. Annotations that do not overlap the text proper or otherwise obscure it are *external*. Such annotations may be in the margins or on separate pages. Annotations such as highlighting and circles around words (which lie over the text proper) or writing which is interlinear we call *within-text*.

Type The *type* of an annotation is a characterization of the physical mark the user-reader leaves on the document. E.g.: textual notes, shapes drawn around the text proper, underlining, asterisks, arrows and other deictic devices.

Following Bradley and Vetch [3, p. 226] we **further** classified marks as casual or meaningful. *Casual marks* were incidental, underlining or highlighting. All others were considered *meaningful*. This distinction follows observations by Charney [4] and others who have found that successful reader-users of hypertexts are so-called *active readers* in Adler and van Doren's [1] sense: they consciously make meaning from text. *Passive readers* read superficially with little or no cognition. Passive readers do not analyse the text.

Readers of both extreme types (passive and active) announce their presence in texts by making marks. Those marks are personal, subjective, and temporal. They are reflective of the readers' most active yet potentially fleeting thoughts. Passive readers' marks are superficial or only used to mark their progress through the text. Active readers sometimes make such marks (i.e., when interrupted during reading or pondering a passage) but most active readers' marks are richer in information.

Type Category The type category is only about the signs user-readers make. We classify the marks users make by type (anchor and content) and function. Function follows in a later category.

Jackson [12, p. 81] noted the 'essential and defining character of the marginal note ... is that it is a responsive kind of writing permanently anchored to preexisting written words'. *Anchor* type annotations serve to call attention or ascribe significance to the part of the document where they are located [20]. Highlighting and underlining are examples of annotation as anchor. *Content* type annotations are notes (drawings, text, etc.) which help reader-users concentrate on parts of the text they find important.

If anchors help the user-reader keep their bearings (in what Dillon [7] calls the information space of the document) then content type annotations orient the reader-users in the argument or draw attention to what they find key. Individual readers have ways of ranking their own marks in part because they generally use a limited repertoire.

Annotations that combine anchor and content types are *compound*s.

Function We classify annotations by their ostensible purpose. Jackson [12, pp. 90,82] observes that not all readers are annotators; the annotator 'acts on the impulse to stop reading for long enough to record a comment'. We categorize the purpose of annotations into 6 classes by the level of engage-

ment following Bloom's taxonomy[6] and Jackson's observations. This order reflects the urgency, and possibly the complexity, of what readers must do to grasp the text:

1. *Interpretive* marks are made when users truly make the text their own: they add some of their own thoughts. An example could be a short note.
2. *Problem-working* marks most often appear near charts or equations, and suggest or record the reader's attempt to understand what is represented or expressed. Definitions are examples.
3. *Tracing progress* may be signaled by the highlighting of lengthy passages, indicating the reader may be overwhelmed by the text, or unable to recognize the relative importance of passages.
4. *Procedural* annotations are intended to draw the user-reader back to parts of the text that require further attention. An asterisk marking a particular sentence could be classified as procedural, for example.
5. *Place-marking and aiding memory* annotations indicate places where the reader signals their presence but not what they are thinking. Highlighting or circling of keywords are examples.
6. *Incidental markings* (e.g. doodles) seem to mark a lack of engagement.

The examples, of course, are the general case as user-readers have their own set of idiosyncratic marks. In our study we found that all of the specific marks correspond to the six categories above.

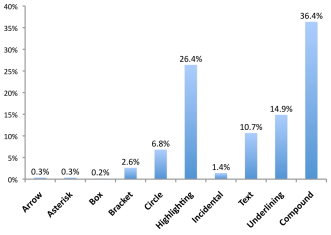
Statistical Method We use a mixed model repeated measures design. Of the fifteen participants, only five volunteered their annotations of both documents. For most we use within-subject ANOVAs which compare multiple measures of each participant's data. Because there were so few participants in both sessions, we computed a between-subjects analysis of 2 groups (one per session). The first *group* was students who participated in only the first session.

3 Results

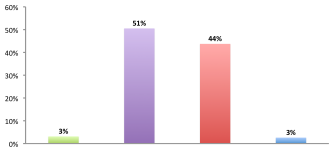
Figure 1 shows that most of the annotations were compounds and that almost all of those include some textual annotation. Table 1 shows the distributions of annotations by count, type category and function. Table 2 shows the distribution of subtypes of marks across all sessions and users; Tab. 2(b) shows the number made by participants who completed both sessions.

Annotation is Idiosyncratic — by Count and Use We found no difference between number of annotations made between sessions by the students who

[6] Bloom's (recently revised) taxonomy is a standard ranking of the levels of learning [11]. The top four of the six levels (namely creating, evaluating, analysing, and applying), suggest the value of the use of natural language. The bottom two (understanding and remembering) do not require the use of language.



(a) All Types of Annotation



(b) Components of Compounds

Fig. 1: Distribution of Types of Annotation

Table 1: Types of Annotations Used ($N = 15$)

(a) Marks Used

Mark	**Mean**	**S.D.**	**Mark**	**Mean**	**S.D.**	**Mark**	**Mean**	**S.D.**
Arrow	7.73	9.00	Asterisk	5.80	10.34	Box	0.07	0.26
Bracket	5.20	6.86	Circle	4.73	6.68	Highlighting	20.27	26.18
Incidental	0.53	0.92	Text	25.47	15.81	Underlining	10.67	13.75

(b) By Category

Category	**Mean**	**S.D.**
Compound	51.53	33.86
Explicit	5.73	6.05
Telegraphic	23.20	16.13

(c) By Function

Function	**Mean**	**S.D.**
Anchor	22.53	14.32
Compound	52.87	37.26
Content	5.07	5.44

participated in both ($t = 0.347, df = 4$). However there were differences between the number of annotations used by the participants taken as a whole ($F(1, 14) = 44.10, p < 0.001, \eta_p^2 = 0.759$). There was a difference between the number of

Table 2: Comparison of Class of Annotations by Session

(a) *Between*-Subjects

	All ($N = 15$)		**First Session** ($N = 7$)		**Second Session** ($N = 8$)	
	Casual	**Meaningful**	**Casual**	**Meaningful**	**Casual**	**Meaningful**
Mean	31.47	49.00	23.00	44.00	30.25	36.63
S.D.	23.18	27.77	16.26	29.70	25.27	21.00

(b) *Within*-Subjects ($N = 5$)

	First Session			**Second Session**		
	All	**Casual**	**Meaningful**	**All**	**Casual**	**Meaningful**
Mean	24.4	13.80	26.80	22.4	20.20	31.80
S.D.	12.4	8.58	15.64	14.6	19.31	20.14

Table 3: p-values of Pairwise Differences in Types of Marks Across Sessions

	Incidental	Problem working	Procedural
Interpretive	< 0.001	0.001	0.001
Place marking	0.001	0.002	0.001

uses of the types of marks ($F(8, 112) = 7.72, p < 0.001, \eta_p^2 = 0.355$) as shown in Fig. 1a. The use of annotations differs by category ($F(2, 28) = 22.599, p < 0.001, \eta_p^2 = 0.759$) and function ($F(2, 28) = 19.741, p < 0.001, \eta_p^2 = 0.585$).

Casual and Meaningful Marks A two-tailed paired within-Ss t-test shows a difference between meaningful and casual marks. More marks were casual than meaningful ($t = -3.335, df = 14, p < 0.005$). Within-Ss analysis showed the number of meaningful versus casual marks differed by participant but not group ($F(1, 13) = 12.568, p = 0.004, \eta_p^2 = 0.492$).

Difference over Time A within-Ss analysis of session and function found no effect of function but a strong session effect ($F(4, 56) = 26.099, p < 0.001, \eta_p^2 = 0.651$). The strong demarcation between functions is evident from Tab. 3.

4 Discussion

Reading is a complex activity [6, p. 4] and yet we found that annotation by scholars for study fits into six basic modes (to use Mangen's [15, p. 404] term). Annotation is clearly idiosyncratic. Marks are personal, subjective, and likely temporal. The use of marks which reveal engagement differs between readers too. Some readers use mostly casual marks and others use many more thoughtful markings. Almost all at some time use what we term 'meaningful' annotation styles that show their engagement with the text.

Variety of Marks for a Few Purposes For every user-reader in our study, the variety of marks was small. As few kinds of marks are used their function should be easily supported by a limited palette of types of mark. Perhaps variation in meaning or significance could be represented by variation in colour, shape, etc.

Importance of Textual Annotation Figure 1 and a statistical analysis not presented here strongly indicate that textual marking is a primary form of annotation. There are other strategies for marking presence but serious engagement can only be through words for '[w]hat a [person] cannot state he does not perfectly know'[7]. Given this it is curious that e-book systems still make it difficult

[7] Quoted (by Gowers [9]) from *The Report of the Departmental Committee on the Teaching of English in England.* H.M. Stationery Office, 1921.

for reader-users to make such notes. Perhaps readers' engagement with e-text would increase if writing — as distinct from keyboard use — were adopted[8].

5 Conclusion and Future Directions

As digital culture eclipses print culture, and as hypertext becomes the dominant medium of publication, the kinds of questions to be asked about annotation and marginal glossing are changing. Documents are not merely available on-line (that is to misunderstand the paradigmatic shift); documents on-line reflect a reconceptualising of text. Notions of permanence attached to the written word are thought of as fetish; palimpsests (literally the residuum of erased text on parchment, metaphorically textual edits thought of as obscured in a final draft) are now marked by digital traces and tags. Accordingly the ways that readers can mark their unique engagement and strategies of annotation are changing. However we must be mindful of what they do currently so that we can support the reasons, if not all of the 'intuitive' or familiar forms [13].

Classification of Annotation Styles of Scholarly User-Readers The classification of marks we developed in our study (§2) accounts for every mark found in the study. A future direction will be to validate or correct that classification by applying it to a wider range of contemporary annotated documents. Retrospective self-analysis or talk-aloud studies are necessary to corroborate our assessment of user-readers purposes in making such marks.

Telegraphic Marks Since the uses of telegraphic (i.e., non-textual) marks are quite limited (although they are certainly idiosyncratic) e-reader tools need only provide a small palette of such marks in several variations.

Textual Marks Annotation captures a person struggling to make meaning and sense. When the user-reader is confronting a new idea, synthesizing it, or capitalizing on it then the engagement must be textual, i.e. with words.

It is clear that support for textual glosses is a necessity for the success of any annotation system for scholars. Precisely how textual annotation should be supported is unclear. Wolfe [26] and Black et al. [2] have shown the importance of simultaneous on-screen presentation of notes that do not obscure the original text. It is not yet clear which potential methods are best. A major distinction in current methods is whether glosses are present when readers view the text proper or if users must act to display the gloss [3, 27].

It is clear that annotation is sufficiently key to the experience of reading that interfaces must be designed to ensure that readers can continue to annotate texts. Knowing why people make annotative marks is more important than knowing

[8] Some have claimed that annotation mediated by keyboards is inferior to annotation with styli because of the parts of the brain that are involved [16], while others conclude that all forms of note-taking require substantial cognitive effort [24].

precisely which marks they make. Of particular importance are textual marks as distinguished from figural marks. To support users' needs digital systems must support the functionality people seek from traditional tools but not necessarily ape users' methods.

Acknowledgments We thank Patricia Oprea for research assistance. We are grateful to Dana Murphy for guidance and assistance with statistical analyses. SSHRC funded this research through an ASU grant administered by Nipissing.

References

1. Adler, M.J., van Doren, C.: How to Read a Book: The Classic Guide to Intelligent Reading. Simon & Schuster, Inc., Toronto (1972)
2. Black, A., Wright, P., Black, D., Norman, K.: Consulting on-line dictionary information while reading. Hypermedia 4(3) (1992)
3. Bradley, J., Vetch, P.: Supporting annotation as a scholarly tool—experiences from the online Chopin Variorum edition. Literary and Linguistic Computing 22(2), 225–241 (2007), http://dx.doi.org/10.1093/llc/fqm001
4. Charney, D.: The effect of hypertext on processes of reading and writing. In: Selfe, C.L., Hilligoss, S. (eds.) Literacy and Computers: The Complications of Teaching and Learning with Technology, chap. 10, pp. 238–263. The Modern Language Association of America, New York, NY (1994)
5. Coleridge, S.T.: Rime of the Ancient Mariner. Dover (1970), [This is the 1817 edition (with marginal gloss); illustrated by Gustave Dore (1870).]
6. Crowder, R.G., Wagner, R.K.: The Psychology of Reading: An Introduction. Oxford University Press, Oxford, UK, second edn. (1992)
7. Dillon, A.: Spatial-semantics: How users derive shape from information space. Journal of the American Society for Information Science 51(6), 521–528 (Apr 2000), http://dx.doi.org/10.1002/(SICI)1097-4571(2000)51:6%3C521::AID-ASI4%3E3.0.CO;2-5
8. Golovchinsky, G.: Reading in the office. In: Proc. of the 2008 ACM workshop on Research advances in large digital book repositories. pp. 21–24. BooksOnline '08, ACM, New York, NY, USA (2008), http://doi.acm.org/10.1145/1458412.1458420
9. Gowers, E., Greenbaum, S., Whitcut, J.: The Complete Plain Words. Penguin (1987)
10. Hauptman, R.: Documentation: A history and critique of attribution, commentary, glosses, marginalia, notes, bibliographies, works-cited lists, and citation indexing and analysis. McFarland & Company, Inc., Jefferson, NC (2008)
11. Hoffmann, R., McGuire, S.: Learning and teaching strategies. American Scientist 98, 378–382 (Sep/Oct 2010), http://www.americanscientist.org/issues/num2/2010/5/learning-and-teaching-strategies/1
12. Jackson, H.J.: Marginalia: Reader's Writing in Books. Yale University Press (2001)
13. Kelly-Bootle, S.: It takes two to intuit. Computer Language pp. 131, 133, 134, 137, 138 (Aug 1989)
14. Lipking, L.: The marginal gloss. Critical Inquiry 3(4), 609–655 (Summer 1977), http://www.jstor.org/stable/1343054
15. Mangen, A.: Hypertext fiction reading: haptics and immersion. Journal of Research in Reading 31, 404–419 (2008)

16. Mangen, A., Velay, J.L.: Digitizing literacy: Reflections on the haptics of writing. In: Zadeh, M.H. (ed.) Advances in Haptics, chap. 20, pp. 385–401. InTech (2010), http://www.intechopen.com/books/show/title/advances-in-haptics
17. Marshall, C.C.: Annotation: from paper books to the digital library. In: Witten, I.H., Akscyn, R.M., Shipman, III, F.M. (eds.) Proceedings of the Third ACM Conference on Digital Libraries. pp. 131–140. ACM Press, Pittsburgh, PA, USA (23 – 26 Jun 1998), http://doi.acm.org/10.1145/263690.263806
18. Marshall, C.C.: Toward an ecology of hypertext annotation. In: Grønbæk, K., Mylanos, E., Shipman, III, F.M. (eds.) Hypertext '98: The Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia. pp. 40–49. ACM SIGLINK and ACM SIGIR, Pittsburgh, PA (20–24 Jun 1998), http://doi.acm.org/10.1145/276627.276632
19. Marshall, C.C.: Reading and writing the electronic book. Synthesis Lectures on Information Concepts, Retrieval, and Services 1(1), 1–185 (2009), http://www.morganclaypool.com/doi/abs/10.2200/S00215ED1V01Y200907ICR009
20. Marshall, C.C., Brush, A.J.B.: Exploring the relationship between personal and public annotations. In: Proc. of the 2004 joint ACM/IEEE conference on Digital libraries. pp. 349–357. Tucson, AZ, USA (07–11 Jun 2004), http://doi.acm.org/10.1145/996350.996432
21. McKendree, J., Reader, W., Hammon, N.: The "homeopathic fallacy" in learning from hypertext. interactions ii(3) (Jul 1995), http://doi.acm.org/10.1145/208666.208687
22. Molotch, H.: Leaps and visions. In: Where Stuff Comes From: How Toasters, Toilets, Cars, Computers and Many Other Things Come To Be As They Are, pp. 68–71. Taylor & Francis (2003)
23. Molotch, H.: The semiotic handle. In: Where Stuff Comes From, pp. 25, 82–84. Taylor & Francis (2003)
24. Piolat, A., Olive, T., Kellogg, R.T.: Cognitive effort during note taking. Applied Cognitive Psychology 19, 291–312 (2005), http://dx.doi.org/10.1002/acp.1086
25. Qayyum, A.: Analysing markings made on e-documents. Canadian Journal of Information & Library Sciences 32(1/2), 35–53 (2008), http://researchoutput.csu.edu.au/R/-?func=dbin-jump-full&object_id=8419&local_base=GEN01-CSU01
26. Wolfe, J.: Annotations and the collaborative digital library: Effects of an aligned annotation interface on student argumentation and reading strategies. International Journal of Computer-Supported Collaborative Learning 3(2), 141–164 (2008), http://dx.doi.org/10.1007/s11412-008-9040-x
27. Wolfe, J.L., Neuwirth, C.M.: From the margins to the center: The future of annotation. Journal of Business and Technical Communication 15, 333–371 (Jul 2001), http://jbt.sagepub.com/content/15/3/333