

## KAIRION: TOWARDS A TECHNOLOGY-BASED PEDAGOGY OF SOURCE USE PRACTICE

Andreas Karatsolis  
Carnegie Mellon University Qatar  
Qatar

### Abstract

Research literature artifacts are not autonomous constructs, but elements of rhetorical processes embedded in specific times and places (Bazerman, 1988; Geisler, 1994; Prior, 1998). Specifically for academic citation practices, Cronin (1984) and Cozzens (1989) argued for a rhetorical perspective of scientific attribution, focusing on the persuasive and performative nature of the practice. This paper reports on the results of textual analyses and discourse-based interviews with 16 researchers from four different disciplines, to argue that source use practices are deeply embedded in rhetorical purpose. Based on this research, the development of a computer-based visualization framework for source use, *Kairion*, aimed at supporting this understanding of academic practices will be presented.

### Background

The role of texts in the development of scientific communities has been the topic of extended discussions in many fields, especially after the 1970s when the myth of the solitary research scientist conducting experiments in the lab or recording data in the field was debunked. Latour and Woolgar (1979) went as far as to suggest that the research labs spend more time and energy in preparing articles, texts that establish their position in relation to other labs, than discovering new knowledge. However, the model of scientific activity as an enterprise chiefly concerned with the improvement of knowledge, as this is developed by individuals (Merton, 1968), still had a very strong appeal. The reason is that in this “republic of science” model, individuals are considered part of meritocratic systems where their rewards or prestige depend on the relative success of the other scientists in the community. Texts, therefore, become the vehicles to provide evidence of continuity and advancement of knowledge as well as innovation.

Such an understanding of the scientific community as an “invisible college” where scientists work and learn in small, self-regulated, merit-based scientific communities (Crane, 1972) has distinct implications about the value of texts: the crucial difference between scientists is not *how* they write, but simply *what* they write. The texts then that the individual scientist produces in his/her community are merely repositories of knowledge, which allowed for the development of the *Science Citation Index* as soon as the technology allowed the creation of large databases. The *Science Citation Index* and consequently the development of the

field of scientometric analysis begin with the assumption that the meaning in scientific texts can be understood by analytically defined indicators (such as co-words or co-citations).

This assumption has been at the heart of attempts to develop both hardware and software which would support the practice of reading and writing from sources. The introduction of Tablet PCs and dedicated reading devices (Sony Reader, Kindle) into the market, as well as a number of annotation systems (Annotator, CiteSeer) and several bibliographic databases (EndNote, Refvis, BibRex) came with the promise of supporting academics access, read, take notes, store and retrieve of information from sources. However, these systems have not gained the popularity of word processors or search databases. The reason seems to be that in most of these systems texts are treated as static objects with clearly identifiable attributes, or *descriptive metadata*, which can be used to store, retrieve and make connections between them in presumably meaningful ways: authors' names, publication, date, subject keywords, quotes, and summaries are all static fields that are already embedded within the document and are even sometimes available for download through a library's database. Even one's notes about a source need to be modified and placed in the context of the project or the argument put forward in each case. It seems that the underlying assumptions about the nature of texts that the developers of electronic tools have would have to be examined more closely.

This rather limited view of texts and their relationship to each other, as it is evidenced by citation practices, has been challenged in the past 25 years by research and theory coming out of citation analysis, discourse analysis and rhetorical theory. In the next three sections I will present the most important articulations of these challenges to such an understanding of texts and citation practices in order to be able to move to the specifics of my own study.

### **Citation Analysis**

In the beginning of the 1980s, the need for a theory of citation became evident in the field of citation analysis through the work of Cozzens (1981) and Cronin (1984). Blaise Cronin in his heavily cited monograph on the citation process in scientific communication (1984) acknowledged that quantities and frequency distributions do not provide an adequate understanding of the nature of citations and called for analyses of both the contexts citations are used and processes by which authors use them. Following Martyn (1965), he proposed that citation is not a unit, but an event, which is very difficult to lay bare as one would have to step into the author's head to understand the functional, social or political motivations behind a single instance of citation. These multiple motivations do not have to be in competition, however — at least not according to Cozzens (1981) who suggested that a theory of citation would be possible if the conversation shifted

away from the traditional sociology of science and included people with different methodological orientations.

### **Discourse Analysis**

A line of research that emerged from the field of discourse analysis is related to the identification of specific patterns in different genres as they emerge from corpus analyses. One such pattern is citation, which is of particular interest for this discussion. The work of Ken Hyland (1999, 2001, 2002) especially stands out here since he has analyzed citation patterns through several measures, including reporting verbs, integral/non-integral citations and self-mention. His conclusions were very interesting for the understanding of disciplinary practices, especially in terms of understanding the knowledge construction process of different groups and the epistemological and social conventions of disciplines. His analyses of self-mention in research articles (2001) also touched upon issues of authorial presence and the construction of identity in academic writing.

Using a similar method to analyze citations from corpora, Thomson and Tribble (2001) examined the difference between integral and non-integral citations in doctoral theses from agricultural economics and agricultural botany. They found differences not only between these seemingly similar disciplines, but also between levels of participation, where novice writers seemed to use a limited range of citation types. However, such studies do not provide any insight as to the reasons why or the motivation behind this limited use of citation types. On the other hand, more ethnographic approaches in the form of one or two case studies of graduate students writing (Connor & Kramer 1998) reveal only a very narrow portion of reality, mostly related to the motivation of these individual students in relation to their discipline or their advisor's demands. Such a qualitative approach, however, has the potential to reveal a much richer picture of citation practices, especially if it is combined with a quantitative understanding of disciplinary methods.

### **Rhetorical Studies**

Research in rhetorical studies has challenged the traditional understanding about texts as static repositories of knowledge and instead proposed seeing them as "virtual objects," objects that do not yet exist except in the mind (Medway, 1996). Geisler (2001) argued that texts play multiple roles in the mind of writers including being a driving motive or a desirable outcome. When texts are mediated by electronic systems, the virtuality of texts becomes even more evident, so the opportunity to develop systems which will make this state transparent is enormous. However, focusing only on descriptive metadata limits to a great extent the potential for seeing them as virtual objects, creating meaningful attributes and connections between them, and eventually providing a context for the use of these sources in writing. These virtual, rhetorical objects transmit information, or

*rhetorical metadata*, which can be equally or more important than descriptive metadata when academics decide which sources to use and how in their work.

This paper has a dual purpose: first it reports on the results of an ethnographic study of academics using citations in order to develop a rhetorical citation model. Then, it presents an example of the way this model can be incorporated into an electronic environment which will support the practices of both old-timers and newcomers in the academia.

## Ethnographic Study

In order to arrive at a rhetorical model of citations, 16 academics from four disciplines (Computer Science, Chemical Engineering, Materials Science Engineering, and Humanities and Social Science) from a large research institute in the Northeast were recruited. They all provided access to at least two of their recently published journal articles to be analyzed for citation patterns and also agreed to a one-hour interview. Each interview included two parts: a typical source use narrative and a discourse-based interview, as it was originally developed by Odell et al. (1983). The typical source use narratives asked participants to tell the story of their reading and writing process when engaged in academic research. These narratives were meant to yield information similar to the typical use case scenarios that designers use to develop electronic systems. Immediately following this interview, they were asked to explain the decisions they had made to at least 15 instances of citation from their own writing in discourse-based interviews. After this model, participants were shown specific instances of source use from their own writing and were asked if using another pattern would have made a difference. All the interviews were then coded automatically for keywords (nouns) associated with either to descriptive or rhetorical metadata (Table 1).

Table 1: Descriptive and Rhetorical Keywords

Descriptive	Rhetorical
Note	Work (noun)
Title	Research
Author	Field
Abstract	Project
Keyword	Argument/Position
Date	Reader
Subject/Topic	Group
Quote	People

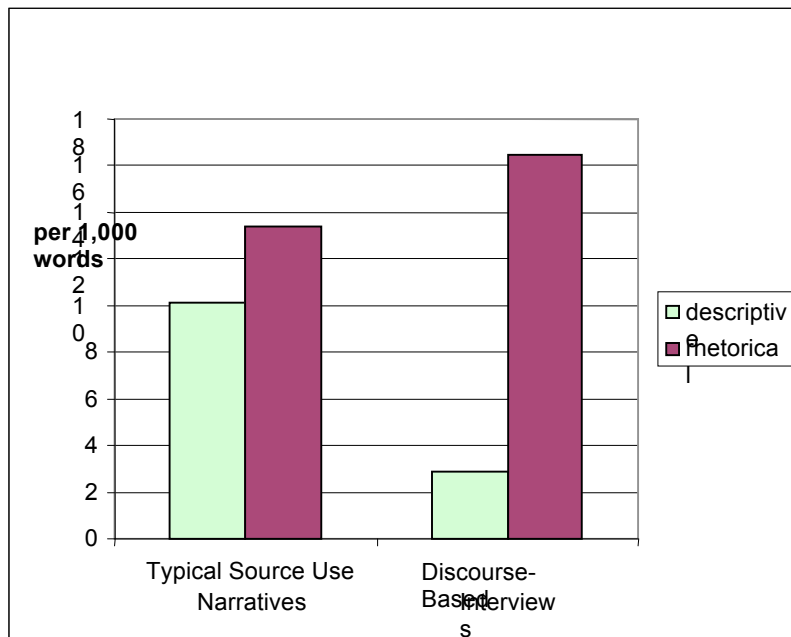
The list of the descriptive nouns was developed based on the fields that electronic systems have modeled for. For example, annotation systems focus on notes and keywords as the most important objects that the users will supposedly interact with, following narratives elicited by users about their usual strategies when reading a source. Bibliographic databases, on the other hand, focus on fields such as title or abstract, which can be downloaded very quickly from a digital library, and can become the objects of a query or search. Finally, reference visualization systems center around dates and topics in order to show the relationship between sources, often over time.

For the nouns associated with the rhetorical perspective, the assumption was that if the descriptive nouns point to a completed, objective and static view of a source, the rhetorical ones should point to the dynamic, subjective and situated perspective. In terms of an activity perspective, the source could be either at the consumption end of the spectrum (as a static object) or at the production end (as a flexible object). In between there are distribution, which divides the objects according to social laws and exchange which further parcels out the already divided shares in accord with individual needs. To capture the activity from the point of production up to the point of consumption, before the source becomes a product which steps outside the rhetorical, nouns were chosen to relate to the basic components of the activity, such as participants (people, group, reader) who do work or research in their field, and develop arguments or positions. In that sense, the rhetorical terms capture the part where a source moves from production to consumption.

The overall results from all participants pointed directly to the rhetorical nature of citation practices. In typical source use narratives, participants used 203 times words associated with descriptive metadata (a rate of 10.1 words per 1,000), while they used words associated with rhetorical metadata 269 times (13.4 words per 1,000). However, in discourse based interviews, the picture changed radically: participants used 726 words associated with rhetorical metadata (16.5 words per 1,000) whereas they used only 126 words associated with descriptive metadata (2.9 words per 1,000).

These results show that academics may be telling a different story about their use of sources when asked to reconstruct it as a narrative than when they are asked to tell the story of specific instances of the way they used sources. While in a typical source use narrative they talk about sources both as static (descriptive terms) and as dynamic (rhetorical terms) objects, in the discourse-based interviews they refer seven times more to rhetorical terms in their effort to justify their practice. This result is the first step to addressing the original assumption about the rhetorical nature of source use.

Figure 1: Overall Comparison of Interview Types



The next step in the analysis was to identify the general categories that participants used in the discourse-based interviews to explain their citation decisions. After analyzing the responses to almost 250 questions, the first pattern that became obvious was that in most instances the decisions to cite a source were directly related to the project itself in relation to the active life of the research community. Readings and consequent writings are always embedded within a context of purpose and audience, as well as a sense of the appropriate moment (or *kairos* in ancient Greek rhetoric) for the field or discipline.

Beyond the dimension of appropriate time, participants identified a big number of critical elements which compelled them to cite sources in their work. As an example, one participant described a decision to include a source like this:

*I think in my field this is probably a fairly well-known statement by now, but you're almost referencing a source because of people like [my co-author] who look at the optimization side of things and aren't going to know this . . .*

Such a statement shows that a consideration for the audience's level of knowledge about the topic was very important, especially since this was interdisciplinary work. If this particular concern was repeated in several of the participants' responses, then I was able to formulate a general evaluative statement of the type: "This source is a good introduction to the problem for new readers/participants."

After combining all the discrete statements developed from the participants' responses, three general categories emerged, which corresponded well with the results of previous research:

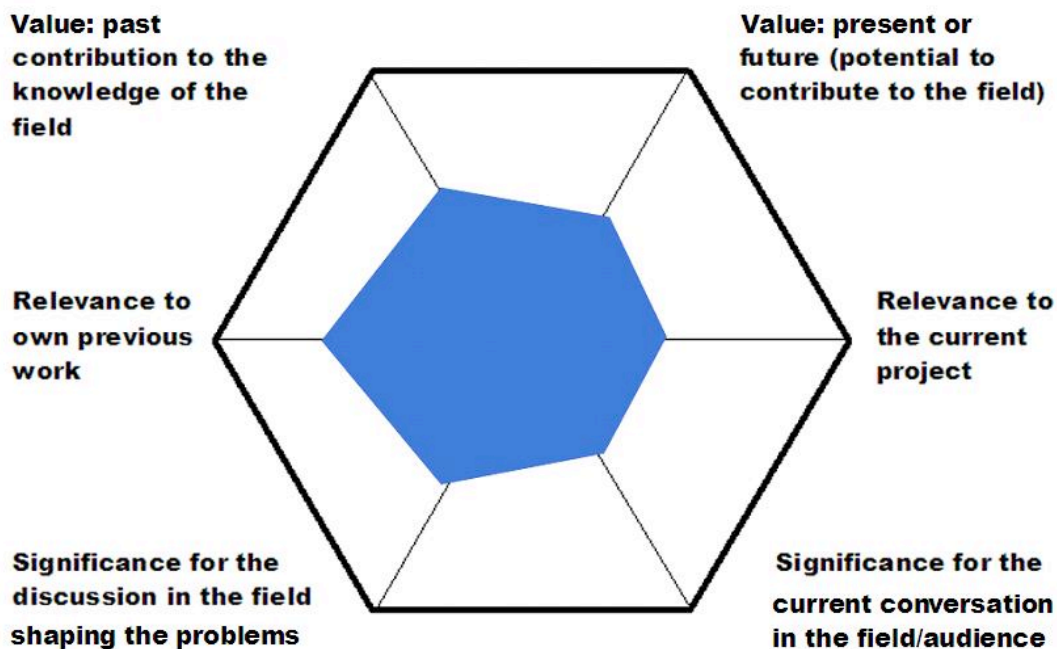
- value (contribution to the knowledge of the field)
- relevance (topical relation to the current argument/project)
- significance (role in shaping the problem for the audience /field)

All these statements in the three categories could then be modeled into an electronic system which would provide not only the descriptive dimension of a source, but also its rhetorical one. In the next section I will present *Kairion*, which stands for "the appropriate, critical" either moment in time or rhetorical consideration of value, relevance or significance which determines when and how a source will be used.

### Development of *Kairion*

These three general categories were utilized to show an aggregate of the ratings, instead of individual data points. This way the schematic of a hexagon would be developed if the source received the full rating for all the questions in all the categories, since there are three categories presented in the dimensions of present and past time. In the following figure (Figure 2) the way the visualization was conceptualized is presented. Both the dimension of time and the three categories can be shown as data points on the hexagon.

Figure 2



This basic model was then developed into an application where the user, after inputting the descriptive metadata of the source, is asked to rate it (1–5) according to the questions associated with each category. There are only four questions for each dimension of the category, for a total of 24 questions. Figure 3 shows the ratings view of the application.

Figure 3: Ratings

**Kairos** Edit Ratings

NEW PROJECT  
OPEN PROJECT  
SAVE PROJECT  
EDIT PROJECT INFORMATION  
EDIT PROJECT STRUCTURE  
EDIT WORKS  
COMPARE WORKS  
EXIT

Epstein, Bruce A.  
Lingo in a Nutshell  
1998

**Value: Past contribution to the knowledge in the field**

1. This source has been heavily cited in your field.  
☒ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

2. At the time of its publication, this source changed the way approached a certain problem or problems.  
☒ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

3. This is one of the first/original/pioneering/classic papers in  
☒ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

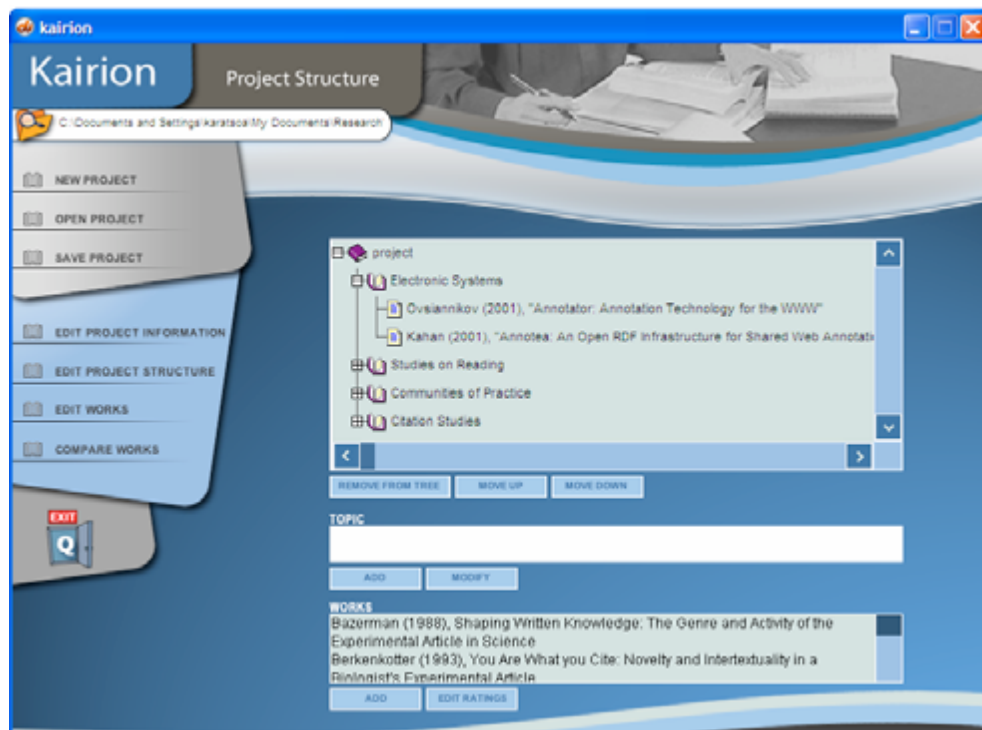
4. This source provides a well-established, comprehensive review/overview of the work in your field.  
☒ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

past present future  
val1 val2  
rel1 rel2  
sig1 sig2

BACK FORWARD

Using Macromedia Director and Flash technology, the hexagon on the upper right-hand corner is constructed as the user is rating the source. This visualization can then be used as a guideline as to what where and why a citation fits in the overall structure of the document or the specific argument the author is making. However, precisely because this information is difficult to recapture at a later time without going through all the visualizations of all the sources again, another view showing the works within the project structure is available (Figure 4). In this view, users can easily connect a certain type of source to their project's outline, for example connect all sources with a high past value (the “classic” sources in the field) to the background section of the document.

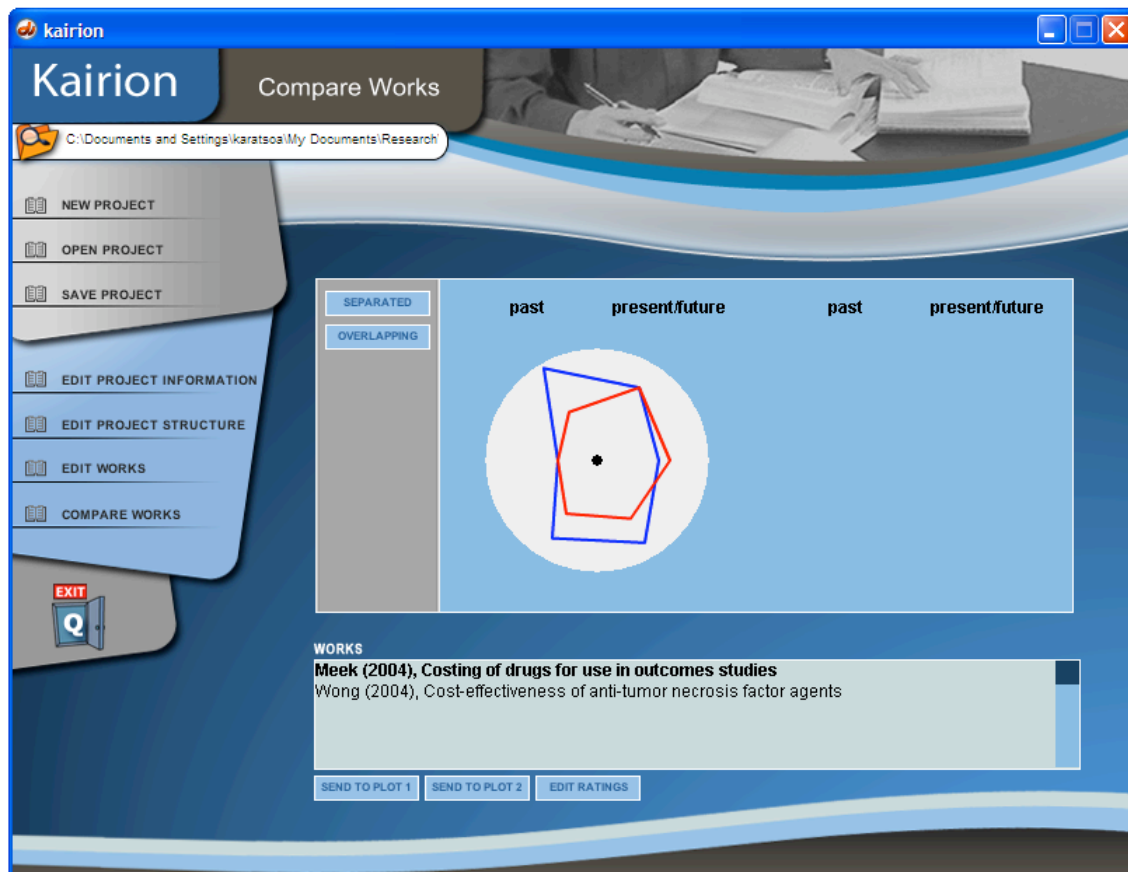
Figure 4



Finally, because very often comparisons between sources can be very useful in determining their relative importance for one of the categories, another view where two visualizations appear either separated or overlapping is available. This view can also show the differences between two readings of the same source, perhaps from an advisor/student pair, or by the same individual for two different projects.

The design and development of *Kairion*, which took much longer than the space it occupies in this brief paper, is meant to provide designers of electronic systems with a way to imagine how rhetorical metadata can be presented. In fact, this use of an interactive visual representation of abstract, non-physically based data to amplify cognition is precisely the definition of an information visualization system, in which the density of useful information must be emphasized (Tufte, 1983). This is exactly the purpose of *Kairion*: to provide academic researchers with a way to visualize, record and share the information which is most useful when working with sources. And this information is certainly not the objective, descriptive metadata that current electronic systems have focused on.

Figure 5



## Conclusion

*Kairion* is only an example of an electronic system which attempts to visualize the rhetorical dimension of sources so that academic practitioners can make more effective decisions about the sources they will use and the way they will use them. Still in prototype phase, *Kairion* would have to be introduced to users and its use has to be analyzed for the duration of a project in order to arrive at conclusions about the accuracy of the three categories and the dimension of time. An item analysis of the individual questions will also have to be performed, ideally with users from different disciplines and stages in their academic career.

The implications of such a system for individual practices are obvious, as academics very often reuse sources for multiple projects or multiple documents coming out of one project. However, for groups of researchers, a common database of rhetorical metadata where the “readings” of the different group members would be transparent has important advantages.

Finally, *Kairion* and similar systems can serve as learning environments where peripheral participants in a field will be able to see in a transparent way how full participants understand the role of a source for their work and for the field. This way, such electronic systems can serve not simply as repositories of information, but as flexible objects aimed at supporting real practices.

### References

- Connor, U., & Kramer, M. (1999). Writing from sources: Case studies of graduate students in business management. In D. D. Belcher & G. Braine (Eds.), *Academic writing in a second language: Essays on research and pedagogy* (pp. 155–182). Norwood, NJ: Ablex.
- Cozzens, S. (1995). U.S. research assessment: Recent developments. *Scientometrics*, 34(3), 351–363.
- Crane, D. (1972). *The invisible college*. Chicago: University of Chicago Press.
- Cronin, B. (1984). *The citation process*. Taylor Graham.
- Geisler, C. (2001). Textual objects: Accounting for the role of texts in the everyday life of complex organizations. *Written Communication*, 18, 296–325.
- Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, 20, 207–226.
- Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. New York: Longman.
- Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics*, 20(3), 341–367.
- Latour, B., & Woolgar, S. (1979). *Laboratory life*. Beverly Hills: Sage.
- Martyn, J. (1965). An examination of citation indexes. *ASLIB Proceedings*, 17(6), 184–196.
- Medway, P. (1996). Virtual and material buildings: Construction and constructivism in architecture and writing. *Written Communication*, 13(4), 473–514.
- Merton, R. K. (1968). The Matthew Effect in science: The reward and communication systems of science are considered. *Science*, 159(3810), 56–63.
- Odell, L., Goswami, D., & Herrington, A. (1983). The discourse-based interview: A procedure for exploring the tacit knowledge of writers in nonacademic settings. In P. Mosenthal, L. Tamor, & S. Walmsley (Eds.), *Research on Writing: Principles and Methods*, (221–235). Longman.
- Thompson, P., & Tribble, C. (2001). Looking at citations: Using corpora in English for academic purposes. *Language, Learning and Technology*, 5, 82–94.
- Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CT: Graphics Press.