

Three Creative Tensions in Document Interpretation Theory Set as Evidence of the Need for a Descriptive Informatics

Joseph T. Tennis

The Information School, Mary Gates Hall, Box 352840,
Seattle, WA 98195 USA <jtennis@u.washington.edu>



Joseph T. Tennis is an Assistant Professor at the Information School of the University of Washington, Associate Member of the Peter Wall Institute for Advanced Study at The University of British Columbia, and Reviews Editor for *Knowledge Organization*. He is also a member of the Dublin Core Usage Board. He received his M.L.S. from Indiana University and the Ph.D. in Information Science from the University of Washington. He works in classification theory, the versioning of classification schemes and thesauri (subject ontogeny), and the comparative discursive analysis of metadata creation and evaluation, both contemporary and historical.

Tennis, Joseph T. *Three Creative Tensions in Document Interpretation Theory Set as Evidence of the Need for a Descriptive Informatics*. *Knowledge Organization*, 36(4), 190-199. 28 references.

ABSTRACT: Describes three tensions in the theoretical literature of indexing: chief sources of evidence indexing, process of indexing (rubrics and methods), and philosophical position of indexing scholarship. Following this exposition, we argue for a change in perspective in Knowledge Organization research. Using the difference between prescriptive and descriptive linguistics as a metaphor, we advocate for a shift to a more descriptive, rather than the customary prescriptive, approach to the theoretical and empirical study of indexing, and by extension Knowledge Organization.

1. Introduction

The word “whom” occupies a strange place in the English language. It is a word seldom heard in speech, and often prescribed in writing. Linguists have observed this phenomenon. They can characterize it as a difference between prescribing how language *should be* used and describing how language *is* used. That is, prescriptive linguistics is the formula for how language should be used and descriptive linguistics is the sum total of how language is used—in the world—not just in text-books. Prescriptivism, starting in the 18th century, can be characterized as having three goals: (1) an aim to codify the principles of language, (2) means of settling disputes over usage, and (3) pointing out common errors thought to be present in language. Descriptivism on the other hand sets as its task to record the facts of linguistic diversity (Crystal 1997).

We can use this distinction as a metaphor to talk about indexing theory and indexing practice. Many of our theoretical discussions of indexing are prescriptive—describing what should be done for improved systems design and implementation. However, we have witnessed an increasing number of studies that move away from this position and are now beginning to describe what is going on in indexing. As a result this model of language proves instructive to our attempt to make sense of diversity in conceptions of and practice of indexing. In this article we follow the descriptivist trend and present three creative tensions in indexing theory. In so doing, we want to also point out how fruitful a more descriptive approach can be to our discipline and to recommendations we may make to systems evaluation.

For our purposes, indexing is a process of interpreting a document for its significant characteristics

using a tool or set of tools in order to represent those characteristics in an information system (Langridge 1989). We are concerned here primarily with the first stage of this process, document interpretation.

Throughout its history, theoretical discussions in indexing—and specifically in the first stage of indexing, document interpretation—have most often been linked with particular controlled vocabularies, classification schemes, or indexes. Many theorists who address the topic of indexing do so in the context of particular systems for classification or indexing (e.g., Ranganathan 1967). If the author is not concerned with a particular system, he or she is interested often in a more general discussion of bibliographic tools (Wilson 1968), domains (Hjørland 2002), or communication processes (Andersen, 2004). As a result, a review of indexing often accounts for this act in concert with bibliographic tools, domains, and communication processes. There are, however, examples that examine the anatomy of indexing, independent of schemes (Mai 2000; Hjørland 1997; Langridge 1989; and Brown 1982). Langridge's *Subject Analysis: Principles and Procedures* (1989) addresses document interpretation without focusing on a particular scheme or other bibliographic tools; however, Langridge imposes an interpretive system much like a classification or indexing system of categories in order to discuss document interpretation. This means that most literature that is pertinent to the first stage of the process, also comments on other stages and components of the process.

In looking at theories of document interpretation, we focus here on the diversity of conceptions of document interpretation theory. By acknowledging this diversity, we believe we can adequately probe the center and survey the perimeter of what has been conceived as a prescriptive discourse. By pointing out these tensions, we can see the need for a broader understanding of this part of Knowledge Organization (KO). Surveying these tensions also allows us to craft a platform for comparative analysis. If all document interpretation theorists approached their work the exact same way, there would be no novelty, and therefore, no need to argue for one way over another way of studying or carrying out such interpretation. However, that is not the case. There is disagreement on how to go about this work. Since there is debate about this difficult problem, identifying conceptual contours and pith of this discourse might lead us toward better systems and practices.

2. Three tensions

A difficult problem like document interpretation elicits a range of thought by theorists, and this fact results in a literature full of creative tension. The point of this paper is to highlight and elucidate that tension. Three key contentions in the indexing literature shape our investigation here. First, there is disagreement about what constitutes the chief source of evidence for a valid interpretation of a document's subject matter. To state it another way, there is disagreement about where the indexer is to "get" a valid interpretation of the significant characteristics of a document. The second contention asks what constitutes a valid process of document interpretation. When is the indexer finished with document interpretation, when is he or she free from constraints on process, and when is he or she tied to a rule, a standard, or a particular best practice? Finally, there are contentions about the relative value of "wrong" philosophical positions and individual research agendas. It seems that document interpretation theory does not allow for multiple research agendas without some critique of wrong ways and the championing of right ways of constructing document interpretation theory. While Hjørland (2002) includes a diverse set of approaches in his domain-analytic approach, he nevertheless also has written that Library and Information Studies (LIS) has had blind alleys in its research (Hjørland 1998). So Hjørland's work is clearly part of the rhetoric about right and wrong document analysis theories.

2.1. *Tension 1: Chief source of evidence in document interpretation*

A major point of contention in document interpretation theory is evidence. What constitutes the chief source of evidence for the valid interpretation of a document's subject matter? Is the chief source of evidence the document itself (Langridge 1989)? Is it the domain's interpretation of the document? Is it the domain and the document together with user and indexer information that governs (Mai 2005)? Is it an individual user's request of a system (Fidel 1994)? Some thinkers take a cognitive approach. They ask whether the valid source of evidence is in the text-processing minds of both indexers and users (Farrow 1991). Others are concerned with how language is deployed to construct meaning—placing the chief source of evidence for document interpretation in a standardized set of discursive practices (Frohmann 1990). The questions of evidence are many. Is the

chief source of evidence a match between words in the document (say, the title) and words in a predefined list (Ranganathan 1967)? Does the chief source depend on your epistemological stance (Hjørland, 1992 and 1997)? Should the ultimate source of indexing consist of a mix of methods and approaches (Bates 1997)? How are all of these approaches related to the representation of the subject in a controlled vocabulary—the end product of document interpretation?

Much work has gone into contemplating this chief source of evidence because, if we were able to identify it, Knowledge Organization, as a discipline, could build better indexing systems and make powerful recommendations for document interpretation practice. As discussed above, it is more common to see discussions of document interpretation linked to a particular scheme than not. However, there is a body of literature that examines this act in and of itself. This body of document interpretation theory is concerned with substantiating the claim for the superiority of one practice over another.

One practice among many roots valid interpretation in the document—the only stable evidence a subject analyst has in his or her hands (Langridge 1989, 5; Mai 2005)—and relates to Hulme's concept of statistical bibliography, which is now commonly understood as literary warrant (Hulme 1923; Svenonius 2001). In this practice, concepts can be recognized from the text at hand. Langridge favours the term *content analysis* to indexing because of its focus on the document and its content (Langridge 1989, 6). Furthermore he separates out use, or potential use, from content (Langridge 1989, 9), and this proves to be an important point of distinction. Yet it is use that nevertheless drives much of document interpretation theory. Mai (2005) adds three more analyses to Langridge's and focuses on use. The domain, the user, and the indexer each serve as use-contexts at work alongside the document (and the indexing language). All of these are necessary for a valid interpretation of a document's significant characteristics. Mai and Langridge's work illustrates the range of approaches in document interpretation theory and stands as a testament to its diversity.

Both Langridge and Mai provide powerfully compelling arguments for their claim about chief sources of evidence for valid interpretation of the subject matter of documents. Given the contemporary LIS emphasis on users, it stands to reason that an explicitly people-centered approach like Mai's will be viewed more favorably than Langridge's document-

centered work. However, on closer examination we may not be able to compare Mai and Langridge in a straightforward manner. It is not a simple matter of document-centered versus domain-centered approaches.

The difference between Mai and Langridge is a difference between the objects of study, focus, and definitions. In Mai's case we are analyzing how a document might be used in a task within a particular domain, seemingly of small enough size that indexers could study the domain and the users and possess enough knowledge about both to make an informed decision about indexing. On the other hand Langridge is analyzing contents of documents based on a mixture of tacit and explicit criteria (not only with one text, the document in hand, but also with the conventions reported in other texts—say philosophical divisions of the universe of knowledge), made manifest in his concept of serving a broad community like a national library (Langridge 1989, 9). The difference between the Mai and Langridge lenses and objects of study is similar to various conceptions of computer science by its theorists. Some might say that computer science is the study of algorithms. And as the science of algorithms, it has nothing to do with interface design. However, another camp might say that computer science is about interface design as well as the study of algorithms. The diversity here is definitional, but it is also about focus and objects of study. Like document interpretation theory, computer science has to make sense of its boundaries. It does so through definitions, detailing focus, and objects of study.

If this is the case, if there is a difference in definitions, focus, and objects of study, then Mai and Langridge may not be studying the same phenomenon. Under the rubric of document interpretation theory, Mai and Langridge may represent two very different camps focused on very different interpretive processes and practices, and researching two very different objects of study. To put it plainly, Mai seems to be studying use analysis, and Langridge seems to be studying content analysis. In making this claim, I am taking a *descriptive* stance rather than a prescriptive stance—where the former acknowledges the diversity in conceptions of indexing and the potential, if not actual, diversity in the reality of practice. The latter prescribes a course of action as correct, above others.

In just this simple set of two types of analysis, a myriad of questions arises. Is content analysis different from use analysis? Do they overlap or are they complementary? What are the assumptions that go

into either? What is the evidence we have to support use of one or the other? Can we use both in information systems? Are both used in the same extant information systems, and, if so, can we observe this phenomenon only if we can acknowledge such diversity? If utility is the arbiter of document interpretation theory, it seems that the utility of either approach is contingent on a complex set of factors—factors at work in practice that become clearer when studied in a rigorous comparative fashion. Answers to these questions surface in a politicized discourse.

Document interpretation theory is a political and politicized research area that seems hostile to variety in philosophical approaches. This may arise from what is seemingly at stake. Document interpretation and indexing research make claims to improve systems and practices. In this milieu, proponents must then claim that there is a single right way to address the analysis. Much of the rhetoric of this body of theory reads as though only one idea holds the ultimate answer to the chief source of evidence (e.g., Hjørland 1997). To the theorists of the domain-centered approach, they have *it*. To the document-centered approach, they have *it*. This point is the scholar's fulcrum—a point used to tip the scales by the gravity of one argument over another. Contrary to both of these views, Wilson claims that no one can have *it* (Wilson 1968, 70-74). He says: "The difficulty in the notion of 'the subject' of a writing is to be located, as it were, in the word 'the' rather than in the word 'subject'" (Wilson 1968, 71 n. 5). In a more descriptive stance, one might extend Wilson's thought on the subject to the chief source of evidence for indexing. In other words, the difficulty in the notion of "the chief source of evidence" for indexing is to be located, as it were, in the word "the." If this is true, a more robust study of conceptions of document interpretation will afford us the critical acumen to choose between, or explicitly combine at will, various conceptions of document interpretation in order to develop more nuanced theory, guide the study of document interpretation in the field, and by extension, better serve our users. Diversity here is seen as an asset not a liability to the full apprehension of theoretical and empirical document interpretation.

The chief source of evidence for valid interpretation of documents is just one place where the diversity among conceptions can be observed. Valid processes of document interpretation are also diverse in nature.

2.2. *Tension 2: Valid process of document interpretation*

There are two types of processes of interest to this review: rubrics and methods. The first is a prescribed set of analysis categories that can be used by the indexer in the interpretive process. The second are suggested techniques the indexer can perform during document interpretation. The variety of the purposes, foci, and definitions of these techniques speaks to a need for a comparative investigation.

2.2.1. *Rubric 1: Request checklist*

Fidel (1994) surveys a number of techniques in indexing. Fidel's concern is with making adequate representations for a user searching a database. One example she posits as promising is request-oriented indexing that uses the checklist method of indexing (Soergel 1975; Fidel 1994). In this case, all the requests that are made of the database are compiled into a list, and if a document would satisfy that request, the document would be indexed under that request. Here analysis is user-centered because it is focused on a user's request of the system. Document interpretation is limited beforehand by requests. Limiting the options for analysis beforehand is not unlike early work on library classifications (Richardson 1964; Ranganathan 1937 and 1967). Yet even if they belong to the same genera, the species are distinct — request checklists do not start with macro-social ideas of requests (i.e., disciplines), but instead focus on individual user requests of a particular database.

2.2.2. *Rubric 2: Kinds of knowledge*

Other thinkers address categories useful to document interpretation (Langridge 1989; Brown 1982; Ranganathan 1967; Coates 1960; and Kaiser 1911; Szostak 2004). These categories serve to separate out kinds of concepts that can be derived from, or ascribed to, the content of documents. Various members of the Classification Research Group also worked with categories. We will focus on the Langridge/Brown categories. Like the Classification Research Group's categories, the Langridge and Brown categories serve as guides to the indexer. Brown's work in 1982 is a programmed textbook that introduces its reader to his technique of document interpretation. This technique is much like Langridge's subsequent work (1989). In fact, Langridge serves as a collaborator for Brown making it difficult to distinguish their individual con-

tributions. Brown's work discusses kinds of concepts a subject analyst might find in an analysis of a document. These kinds of knowledge are: Discipline, Phenomena Studied, Forms of Presentation (including language, whether or not the document is an encyclopedia or dictionary, whether it is a document for beginners or not, among others), and Physical Form. The Langridge/Brown categories are used alongside Ranganathan's Personality, Matter, Energy, Space, and Time (PMEST)—the fundamental categories (Ranganathan 1967). The purpose of the Langridge/Brown categories is to hone the interpretation of the indexer, thereby making it more clear and faithful to the kinds of concepts in the universe of knowledge

In focusing on a faithful representation, the purpose of the Langridge/Brown work is to construct a speciology of writings—to describe the texts as a species of writing. Thus, an *Encyclopedia of the History of Chemistry* contains three kinds of knowledge according to the Langridge/Brown approach to document interpretation. Encyclopedia is a FORM CONCEPT, History is a DISCIPLINE, and Chemistry is the PHENOMENON. Chemistry is what this document is about. Encyclopedia and History tell the subject analyst what the document is—a reference work and a work of historical research. These latter two kinds do not tell the subject analyst what the document is about. It is possible that this general interpretation is not faithful to a domain-analytic view of indexing (Mai 2005) (in this case, there are no data present to substantiate a user-informed set of categories, though it is possible to construct). To that end, Brown acknowledges that particular domains will refine these categories to fit their needs (Brown 1982 n. 116). However, this begs a further investigation, related to our first concern above. From what evidence do these categories derive? Are they tested? How would Knowledge Organization test these categories? These questions are left unanswered in these texts.

Both rubrics discussed above are based on attributes and assumptions that must be better understood. Taking account of these assumptions, we can answer the above questions. This position is a descriptive position. The next section outlines methods as a process of document interpretation.

2.2.3. Method 1: Wilson's four methods

In his foundational work on bibliographic control, Wilson identifies four methods that can be used to say what a particular writing is about. They are: (1)

Purpose Way, (2) Figure-Ground Way, (3) Objective Way, and (4) the Appeal to Unity (or the Appeal to Rules of Selection and Rejection). These four methods are what Langridge would call content analysis. We are primarily, if not solely, concerned with the document.

The first way, the Purposive Way (Wilson 1968, 78-81), seeks the author's intention. If, the argument goes, the analyst finds the author's intention; he or she knows what the document is about. The Figure-Ground Way (Wilson 1968, 81-83) uses a picture metaphor to place one *Cast member* (a topic discussed in the document) at the foreground. It occupies the most space in the picture. This foregrounded cast member is what the document is about. The Objective Way counts references to items addressed in the document (Wilson 1968, 83-86) and is reminiscent of Hulme's statistical bibliography (Hulme 1923), the foundations of literary warrant and bibliometrics (Pritchard 1969). The item that gets the most counts is what the document is about. Wilson's final method is the Appeal to Unity (or the Appeal to Rules of Selection and Rejection) (Wilson 1968, 86-88). In this case the indexer, like the writer of the document, selects and rejects what is to be included in the text. The indexer then makes some unifying statement given the items left over from the analysis of selection and rejection.

Each of these methods has its problems. It is often hard to discern the author's intent. Likewise, it is also quite different to say what the author intended to do as opposed to what he or she actually did. If it is possible to identify both, we then have to ask, which of them is the document about? Likewise, the cast members and references to items each vie for the indexer's attention and upon interpretation are placed in the foreground, counted, or selected and made into a unifying statement. However, as Wilson points out: "[W]e cannot expect to find one absolutely precise description of one thing which is *the* description of *the* subject, all others being mere approximations to that one description, or being descriptions of what is not the subject. The uniqueness implied in our constant talk of *the* subject is non-existent," (Wilson 1968, 90; emphasis added).

2.2.4. Method 2: Grammatical model

Svenonius (1994) offers a grammatical model of the definition of subject, much like Wilson's Counting References method, with a different epistemic stance. This is, as she claims, a positivistic approach to

the definition of subject. In this model, the subject is a proposition, e.g., snow is white. The more that is said about snow, in a document, the more the document is about snow. As writers produce more documents that contain statements about snow, snow becomes a subject of study, and hence is analyzed as such (Svenonius 2001, 47). However, the act of indexing is not clarified by Svenonius. It seems she believes that the subject analyst need only read sentences to identify what is being proposed and that is the subject (Svenonius 2001, 47-48). This is in direct opposition to Wilson's idea that we do not know what we need to know to understand a sentence (Wilson 1968, 77). So this method, compounded with Wilson's arguments, serves as a point needing clarification via comparison with other methods in a comparative investigation.

2.2.5. *Processes and conceptions*

As we can see from the foregoing, theories that recommend different processes of indexing, approach the process from different conceptions of that phenomenon. All are contingent theories. Even if it is often not explicit to the recommender, various processes are based on various assumptions about the object of study, the act of interpretation, the purpose of interpretation, and the goal of interpretation. Here again, Langridge's conception is helpful. We see at work in the sentence above the five attributes of his conception of indexing: (1) analysis, (2) documents, (3) significant characteristics, (4) representation; and (5) information system. Langridge's conception can then be used across multiple contexts. In this particular case, the purpose of the interpretation is linked to representation and to the information system, and also influences what are considered significant characteristics. Another conception of indexing may prioritize the purpose of the interpretation over the goal of interpretation be prioritize matching user requests to terms in the system over creating well-formed and concise index entries. However, the processes of indexing, as addressed in the theoretical literature may define the process in exactly the same way. The expressed varieties of the indexing process require further analysis. A comparative analysis can investigate varieties of processes discussed in the theoretical literature. Conceptions of indexing rely on an epistemic stance and a value system affiliated with it. What follows from such positions is a discussion that often unfolds as critique.

2.3. *Tension 3: Critique of philosophy or a critique of research agenda?*

In his 1990 article, Frohmann makes a valuable argument for increasing the diversity of indexing research. Frohmann's focus is on encouraging indexing researchers to engage in a research agenda based on constructs of meaning, discursive studies of information retrieval, and a philosophical approach to indexing research informed by the later philosophy of Ludwig Wittgenstein (1953). According to Frohmann, an agenda shaped by these matters will serve indexing research much better than studying the human mind.

Frohmann (1990) makes the claim that, by Wittgenstein's lights, indexing research should not be concerned with processes of the mind. Rather, he makes a compelling case for study of the more social and discursive aspects of indexing. However, Frohmann makes a significant leap—a leap that calls into question the nature of his critique. Frohmann's leap is from an exposition of the goals, assumptions, purposes, and products of the mentalists' research agenda to a list of the benefits of a Wittgensteinian research agenda of indexing. This move is a leap because he provides no compelling argument to refute foundational aspects of mentalism, its definitions, assumptions, purposes, and products. Nor does the Wittgenstein agenda completely supplant the mentalist interests in indexing in terms of definitions, assumptions, purposes, and products. In short Frohmann seems to be comparing apples to oranges. He is comparing a research agenda of the mind against a research agenda of discursive activity.

Instead of taking an ecumenical path, he claims that diversity limits indexing research. His contribution to the broader scope of indexing research agendas is valuable. It is an example of the diversity of conceptions of indexing. However, he does not successfully supplant mentalism because he does not explicate all of its flaws, nor does he take it for what it claims to be. This complicates the reader's view of his work. How is the reader to understand his rhetoric?

It seems that Frohmann's reader can ask, is his critique a critique of the foundational philosophy of mentalism? If it were such a critique, we would expect to see a logical argument for why it fails as a philosophical foundation. (Wittgenstein's later work (1953) may serve this purpose for Frohmann; but, it should be acknowledged that Wittgenstein's work is not sacrosanct. Wittgenstein's linguistic anxiety is a symptom of Modernism, according to Latour (1993),

and if we follow Latour we are no longer in a condition of Modernity with its symptoms. In fact, to Latour we have never been modern.) Or, is Frohmann's critique one of someone else's research agenda? Is Frohmann critiquing definitions, assumptions, purposes, and products? At what point is Frohmann making a claim about valid knowledge of indexing, and at what point is he saying "I do not believe in this perspective, therefore indexing, as a whole, should not engage in this research?" It is unclear whether Frohmann has adequately critiqued definitions and assumptions of mentalism, and further, we can only see that Frohmann is offering a rationale for studying indexing in another way without substantially refuting mentalism. What then is the evidence for these beliefs?

Evidence that Frohmann has not considered definitions and assumptions of mentalism can be found in his opening paragraphs. First, Frohmann defines what he calls indexing as having two distinct operations. "The first involves either the implicit or explicit representation of a document by an indexing phrase. The second involves the translation in the terms of the indexing phrase into the lexicon of a controlled indexing vocabulary" (Frohmann 1990, 82). Here his assumptions illustrate his misunderstanding of the mentalist's work. The second is his excerpt from Beghtol (1986). Beghtol states explicitly in her work what she is studying; and, it is not what Frohmann suggests indexing should study. We will discuss Frohmann's definition first, and then take up his excerpt from Beghtol (1986).

In the definition of indexing, Frohmann has ignored a number of factors that are later considered by Mai (2001, 593-595). He ignores the initial interpretation process of the document and all the factors that go into interpretation before a subject analyst is able to represent the document by an indexing phrase. Mai (2001) calls this act—neglected in Frohmann's definition—document analysis. It is this act of document analysis that is crucial to the subsequent steps that Frohmann uses as his definition of indexing. And though Frohmann is not using document analysis as part of his definition of indexing, it is precisely this act (and subsequent acts in indexing) that Beghtol (1986) studies. And this is evident from Frohmann's excerpt of her work. Frohmann illustrates his misunderstanding of Beghtol's study by excerpting from her work a block of text that tells the reader what she is studying. Frohmann does not acknowledge all of the levels she mentions, some of which would belong in Frohmann's more social and

discursive research agenda. Beghtol says, and Frohmann excerpts (Beghtol 1986, 90, 92 quoted in Frohmann 1990, 83-84):

During the act of reading a text the reader notices the presentation of each sentence, automatically transforms its surface verbal structures into its deep conceptual propositions and establishes an understanding of the logical relationships between the words and the sentences of the text At the same time, the reader engages in a global, textual or macro-level analysis of the text in order to arrive at the overall understanding of the aboutness and meaning of the complete text as a whole These cognitive actions of compressing text in order to generate a semantically accurate statement of discourse aboutness are, according to Van Dijk, governed by macro-rules One may say that the subject of a document is the highest specific macroproposition that is produced and can be expressed by a reader during cognitive reduction of a text by microanalysis Van Dijk has formally described and analysed a cognitive process that can be assumed to operate during the aboutness analysis of a text for the purpose of classifying it by means of a particular classification system.

From this excerpt, it seems that the reader of a text, when transforming surface verbal structures into deep conceptual propositions or engaging in global, textual or macro-level analysis of the text is engaged in a social or discursive action, at least to some degree. This is clear from the links to language, common to social groups and the classification systems—a shared formal language standard. Even if macro-level analysis is only a small part of what Beghtol is studying, there is no evidence to support the claim that Beghtol would not welcome studies of macro-level analysis, or, by extension, discursive analysis.

The problem becomes clear that they are talking about different processes given Frohmann's definition of indexing, and given Beghtol's focus, as excerpted in Frohmann. Frohmann is trying to make a point that, by his definition of indexing, researchers of indexing should adopt a Wittgenstein-influenced approach. However, Beghtol is not working on exactly the same problem as Frohmann. She is concerned with summarization and text condensation, as mentioned above. She is concerned with the "cognitive process of classifying documents" (Beghtol 1986, 84). She is not concerned solely with the im-

PLICIT or explicit representation of a document by an indexing phrase (Frohmann 1990, 82). Her project is broader and deeper than that. It is deeper because of its focus on a theory for an experimental design, and it is broader because it includes social and discursive practices (Beghtol 1986, 85 and 98). Beghtol's 1986 article is a particular type of research that if KO were concerned with a comparative stance, must be taken on its own merits. It is just as valuable to the diversity of conceptions of indexing as Frohmann's. Therefore it is valuable for our deeper understanding of the conceptions of indexing. Perhaps, Frohmann feels the same way, even with his critique.

The point of Frohmann's critique comes at the end of this article (1990, 94):

[Mentalism's] danger lies not in building systems on flawed foundations, but in its power to deflect attention from theoretical problems central to the development of effective information retrieval systems. I suggest that mentalism conceals fruitful directions of enquiry.

Frohmann's critique is not a critique of a philosophy of the particulars of Beghtol's and others' research; it is not a critique of mentalism. Frohmann's critique is on general research focus. Frohmann protests against how researchers in indexing theory do not study indexing from a more social and discursive vantage point.

Making sense of Frohmann's point and his arguments is not easy. Layers of rhetoric, implicit as well as explicit, hidden as well as overt, shape his argument. The same can be said of Beghtol's work, at least to some degree. But it is precisely this complexity that needs to be studied. Are indeed Frohmann and Beghtol researching at cross-purposes? Is Frohmann's critique compelling to the point that all researchers in indexing will abandon mentalist paradigms? What components of Frohmann's critique are similar to Beghtol's conceptions of indexing, what are different? A comparative approach can elucidate the crossover between the two, as well as the disparities.

Frohmann's article critiques indexing research in LIS. Critiques like these champion one perspective. A more thorough understanding of the differences and diversities in indexing theory is needed to understand what is at work in critiques of philosophical foundations. Such an understanding can be garnered from a descriptive investigation of conceptions of indexing.

3. Need for a study of conceptions of indexing, and a need for a descriptive informatics

In summary, three key contentions in indexing literature make a compelling case for a comparative re-examination of conceptions of indexing, or, in other words, a descriptive approach. First there is disagreement about the chief source of evidence for the subject analyst. Is the chief source of evidence the document, the user, the domain, the request, all of these? The literature is not unified on this issue. Secondly, what constitutes the valid process of indexing is contested. When the subject analyst conducts indexing, what processes does he or she go through? When is he or she finished with the process? This is a complex issue and requires a closer examination. Finally, there is a debate about the propriety of one philosophical tradition over another in indexing research. In this third point of contention there is no clear comparison between the unity and diversity of the approaches that each philosophical perspective takes. Furthermore, there appears to be hostility toward particular approaches to indexing research (e.g., Frohmann's attack on mentalism). The politics and substance of these perspectives can and should be further explored. These three points of contention point to the need for a comparative study, following a descriptive approach, of conceptions of indexing.

It is assumed that these points of contention surface because the design, implementation, and evaluation of systems and practices is at stake—that one way is the right way and others will lead to poorly constructed indexes and catalogues. A descriptive approach can provide such a comparative study of the conceptions of indexing. Opinions range concerning the chief sources of evidence for valid interpretation of documents. They can be compared and contrasted. There are a host of processes that can be considered valid, well beyond what was addressed here, each with its accompanying philosophical bent. Each conception of the process can be analyzed in order to flesh out what assumptions, definitions, and objects of study are at work in their particular conception of indexing. It has also been shown that critiques can be seen as offering an alternative research agenda, and, by extension, adding more diverse accounts of conceptions of indexing. Such diversity begs analysis and synthesis. A descriptive investigation offers the mechanisms for a comparative analysis.

The range of factors in indexing influences our conception of it as a professional act. We also talk about this particular practice in a general way—a way that

may be generalizable to other procedures. Beghtol (2003) describes such a study. Her work on naïve and professional classification is a descriptive study—one that takes a general definition of phenomena and applies it to study a broader universe of the classificatory act. In doing this Beghtol is speciating types of classification. There is a need for such speciation because we have a diverse universe of KO acts and systems—ontologies and social tagging constitute two major examples alongside user-centered and document-centered indexing, and professional and naïve classification.

An approach of descriptive investigations—a descriptive informatics—would illuminate qualities and the nature of the panoply of practice that constitute the wide and diverse universe of contemporary (and historical) KO. We are borrowing the term *descriptive* from linguistics. Descriptive Linguistics describes how language is spoken. It accounts for the variation and the redundancies of the language through comparative study—removed from the prescriptive conception of how language should be spoken.

A prescriptive stance to language is an educational function. It provides the rules for how language should be spoken. Linguistic prescription, would tell the English speaker when to use “who” and when to use “whom” in a sentence. This is like the discussion above that advocates for a user- vs. a document-centered approach to indexing. Linguistics descriptions tells us who, when, how, and where “whom” is used instead of “who.” Likewise, a descriptive approach in KO, and specifically a descriptive approach to indexing, would say who, when, how, and where user-centered indexing is employed, what epistemic stances play out in ontologies, folksonomies, and library catalogues. And a descriptive approach would follow Beghtol’s work and compare structures and process of naïve and professional classification.

The definition and preliminary anatomy of indexing research (evidence, process, and philosophical critique) outlined above might aid in the common communication of a descriptive approach. It might be convenient and meaningful to say the same things about the process, significant characteristics, representations, and information systems used in different types of indexing practices—naïve and professional, done by ontology engineers or social taggers. We feel there is much more to learn about our conceptions of KO from such a comparative agenda. And a descriptive informatics—one that takes descriptivist approach and compares extant methods, techniques, tools, and reflections on knowledge organization will aid us in understanding more.

The phenomenon of indexing is complex. Our theories of document interpretation have showed us just a few of the factors that influence our understanding of the act and its contingencies. However, our theoretical work has most often spent time prescribing a theoretical position that weights one “correct” conceptualization. This is a reasonable position if KO were only about prescribing a single way of indexing. However, such an approach obfuscates the diversity present in the theoretical literature in terms of the phenomenon of indexing and in the complex suite of factors that shape it. A descriptive informatics would bring that diversity to the front—making sense of the tensions present in scholarship and explaining the explosion of types and instances of similar yet different indexing systems.

References

- Andersen, Jack. 2004. *Analyzing the role of knowledge organization in scholarly communication: An inquiry into the intellectual foundation of knowledge organization*. Copenhagen: Department of Information Studies, Royal School of Library and Information Science.
- Beghtol, Clare. 2003. Classification for information retrieval and classification for knowledge discovery: Relationships between professional and naive classifications. *Knowledge organization* 30: 64-73.
- Beghtol, Clare. 1986. Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. In *Journal of documentation* 42: 84-113.
- Brown, Alan George. 1982. *An introduction to subject indexing*. London: Bingley.
- Crystal, David. 1997. *The Cambridge encyclopedia of language*. Second edition. Cambridge: Cambridge University Press.
- Farrow, John F. 1991. A cognitive process model of document indexing. *Journal of documentation* 47: 149-66.
- Fidel, Raya. 1994. User-centered indexing. *Journal of the American Society for Information Science* 45: 572-8.
- Fidel, Raya. 2000. The user-centered approach: How we got here. In Wheeler, William J. ed., *Saving the time of the library user through subject access innovation*. Champaign, IL: Graduate School of Library and Information Science, pp. 79-99.

- Frohmann, Bernd. 1990. Rules of indexing: A critique of mentalism in information retrieval theory. *Journal of documentation* 46: 81-101.
- Hjørland, Birger. 1992. The concept of subject in Information Science. *Journal of documentation* 48: 172-200.
- Hjørland, Birger. 1998. Theory and metatheory of information science: A new interpretation. *Journal of documentation* 54: 606-21.
- Hjørland, Birger. 2002. Domain analysis in information science: eleven approaches – traditional as well as innovative. *Journal of documentation* 58: 522-62.
- Hulme, Edward Wyndham. 1923. *Statistical bibliography in relation to the growth of modern civilization*. London: Butler and Tanner.
- Kaiser, James. 1911. *Systematic indexing*. London: Pitman.
- Langridge, Derek Wilton. 1989. *Subject analysis: principles and procedures*. London: Bowker-Saur.
- Latour, Bruno. 1993. *We have never been modern*, trans. Catherine Porter. Cambridge, MA: Harvard University Press.
- Mai, Jens-Erik. 2005. Analysis in indexing: document and domain centered approaches. *Information processing and management* 41: 599-611.
- Metcalf, John Wallace. 1957. *Information indexing and subject cataloging: alphabetical, classified, coordinate, mechanical*. New York: Scarecrow Press.
- Pritchard, Alan. 1969. Statistical bibliography or bibliometrics. *Journal of documentation* 25: 348-9.
- Ranganathan, S. R. 1937. *Prolegomena to library classification*. Madras: Madras Library Association.
- Ranganathan, S. R. 1967. *Prolegomena to library classification*. 3rd ed. Bombay, India: Asia House Publishing.
- Richardson, Ernest Cushing. 1964. *Classification: theoretical and practical*. Hamden, CT: Shoe String Press.
- Soergel, Dagobert. 1975. Theoretical problems of thesaurus construction with particular reference to concept formation. In Petoefi, R., ed., *Fachsprache-Umgangssprache*. Kronberg, Germany: Scriptor, pp. 355-81.
- Svenonius, Elaine. 1994. Access to nonbook materials: the limits of subject indexing for visual and aural languages. *Journal of the American Society of Information Science* 45: 600-6.
- Svenonius, Elaine. 2001. *Intellectual foundation of information organization*. Cambridge: MIT Press.
- Szostak, Rick. 2004. *Classifying science*. Springer.
- Wilson, Patrick. 1968. *Two kinds of power: an essay in bibliographical control*. Berkeley, CA: University of California Press.
- Wittgenstein, Ludwig. 1953. *Philosophical investigations*. Oxford: Basil Blackwell.