D. Grant Campbell—University of Western Ontario

Tensions Between Language and Discourse in North American Knowledge Organization

Abstract

This paper uses Paul Ricoeur's distinction between language and discourse to help define a North American research agenda in knowledge organization. Ricoeur's concept of discourse as a set of utterances, defined within multiple disciplines and domains, and reducible, not to the word but to the sentence, provides three useful tools for defining our research. First, it enables us to recognize the important contribution of numerous studies that focus on acts of organization, rather than on standards or tools of organization. Second, it gives us a harmonious paradigm that helps us reconcile the competing demands of interoperability, based on widely-used tools and techniques of library science, and domain integrity, based on user warrant and an understanding of local context. Finally, it resonates with the current economic, political and social climate in which our information systems work, particularly the competing calls for protectionism and globalization.

Introduction

As the North American knowledge organization community gathers to assess its history and its future, I would like to explore a persistent tension between universal and local perspectives in knowledge organization practice and research. I will use the theory of language posited by Paul Ricoeur (1977) to suggest that our traditional tendency to view subject tools as "languages" is a valid and useful one. But I will also suggest that Ricoeur's attempt to expand views of language beyond the purely semiotic realm provide a useful guide for describing and defining our own research. In particular, I will suggest that Ricoeur's distinction between semiotics and semantics—between language as a self-enclosed, self-referential system and discourse as an open, multi-disciplinary process of speech events—suggests ways of negotiating the ongoing tension in knowledge organizaton between the local and the universal in an economic, technological and social environment that struggles to reconcile the regional and the global in human life.

Domain knowledge has long been recognized as an essential part of information organization, particularly in the understanding of literary and user warrant (Lancaster 1986) and the provision of services for special libraries (Foskett 1966, xiii). However, the very act of investigating context can be seen as the application of professional tools that are distinct from that context. In 1999, Marcia Bates argued that representing information requires a specific skill set which is distinct from actually "knowing" the information, and defended the distinction by comparing information professionals to professional actors:

> We take it for granted that when we see a film or television program like "ER" ("Emergency Room"), that it is actors who portray the physicians, because that is the way it has always been done. ... In like manner, *representing* information—whether you are indexing or formulating a search strategy or helping someone articulate what they want to find—is different from *knowing* the information. ... Creating databases and catalogs involves creating representations of forms of information. (1999, 1045)

This familiar tension between specific subject knowledge and general representation strategies has acquired a fresh resonance in 2009, given its resemblance to the tensions between the global and the local in North American economies, and the conflicting arguments for protectionism and for free trade in the wake of the recent world recession. Like North American policy makers, North American information professionals must negotiate conflicting demands and allegiances: to specific user communities on the one hand, whether defined by place or by fields of knowledge, and to practices, tools and standards of information organization embraced by national and international library communities. This paper will address one specific aspect of this challenge: how can our information systems establish optimal levels of interoperability between one system and another: interoperability that enables community members to gain access to global information resources, while still preserving the data structures, information models and community affordances that make their systems a manifestation of their distinct cultures and communities? And how can the North American knowledge organization community establish research that supports information professionals as they work to reconcile global and local perspectives?

Subject Access Systems as Languages

A partial answer to those questions can be found in the practice of characterizing thesauri and classification systems as specialized languages. Elaine Svenonius (2000), in her ambitious history of information organization, argues that information systems draw heavily on language theory. The linguistic dimension of subject access tools is evident in such terms as "controlled vocabulary" and "thesaurus," and the use of these tools involves the inherently linguistic concepts of vocabulary and syntax: the establishment of authorized terms and their inter-relationships, and the use of a standardized syntax for concatenating terms together to form classification numbers or precoordinate subject headings. Other theorists such as Blair (1990) have explored the similarities between information description on the one hand and semiotics and linguistic analysis on the other. Others have linked subject access to post-modernism (Mai 1999) and post-structuralism (Campbell 2008). The use of linguistics as a paradigm for subject access systems has placed linguistic theory, with its strengths and its limitations, close to the heart of subject analysis. For this reason, Paul Ricoeur's analysis of this linguistic tradition provides a means of enabling us to recognize some of the limitations of classical linguistics, and to prevent those limitations from hindering the growth of an active and vital North American research agenda.

The Linguistic Approach: Saussure and Ricoeur

In his highly-influential *Course on General Linguistics* (1916), Ferdinand de Saussure articulated numerous propositions that have since been widely adopted, not as empirical evidence of how language works, but as metaphors for how language-like systems such as subject tools work. First, a language can be considered in two separate ways. For Saussure, linguistics is primarily concerned with language as *langue*, which he defines as the entire system of linguistic units and the code that relates them together: he distinguishes this "self-

contained whole" (1985, 29) from the "executive" domain of specific "speech acts," which he terms *parole* (32). This distinction, for Saussure, separates the individual from the social and the incidental from the essential: "Language is not a function of the speaker; it is a product that is passively assimilated by the individual. ... Speaking, on the contrary, is an individual act. It is wilful and intellectual" (33).

Second, Saussure argued that the "sign" was a combination of concept and sound-image, or of "signified" and "signifier." Furthermore, he argued that the sign is intrinsically arbitrary, and derives its meaning, not from any innate relationship between the word and the thing, but from the differences between words in the *langue* system. As Ricoeur points out, the sign rests on a play of differences: "in such a system no entity belonging to the structure of the system has a meaning of its own; the meaning of a word, for example, results from its opposition to the other lexical units of the same system" (1976, 5).

In the first of a series of lectures delivered at Texas Christian University in 1973, Paul Ricoeur took issue with Saussure's linguistics: not with its importance or validity, but with the ensuing impact Saussure's treatment had on linguistic theory that came after, and its effect on questions of words and truth that went back at least to Plato (Ricoeur 1976, 1). For Ricoeur, Saussure's emphasis on the importance of *langue* had blinded theorists to the importance of *parole*. *Langue*, Ricoeur argued, had the advantage of being self-contained and theoretically finite, and generally studied within a single discipline: that of linguistics. *Parole*, on the other hand, is theoretically infinite and, as an event, can take place and be studied within a variety of disciplines. Ricoeur suggests an alternative, two-dimensional approach to language which rests on two irreducible entities: the "sign," which emerges through the *langue*, and the "sentence," which emerges from the speech-acts that constitute *parole*, and which he terms "discourse." For Ricoeur, the sentence is a basic and intrinsic unit related to semantics, rather than semiotics. While a sentence is composed of words, its propositional content cannot be reduced to its words: it remains a union, however succinct, of a noun and a verb (Ricoeur 1976, 10).

When we take Ricoeur's two-part transformation of Saussure's linguistic theory and transfer it into the realm of subject tools and knowledge organization, a suggestive similarity presents itself. Language deals with signs and sentences: with semiotic units of a closed, self-referential system, and with semantic units of spoken and written discourse, embedded in their social, cultural, political and economic context. Similarly, Beghtol draws on Robert Fairthorne's important distinction between two dimensions of "aboutness": the subject content of a document that must be rendered using the signifying system of the subject tool: "Extensional aboutness, in Fairthorne's terms, is the inherent subject of the document; intensional aboutness is the reason or purpose for which it has been acquired by a library or requested by a user" (Beghtol 1986, 84). This relationship has always been fraught with tension, as subject cataloguers attempt to negotiate the anticipated needs of the user with a perceived "essential" content that could be put to a variety of uses. Some work on the assumption that "a document has an intrinsic subject, an 'aboutness', that is at least to some extent independent of the temporary usage to which an individual might put one or more of its meanings" (Beghtol 1986, 85). Others operationalize aboutness as an estimate

about probable search behaviour (Maron 1977, 38), while still others argue that we should be moving from a document-centric (and presumably extensional) approach to a domain-centered, contextual, and implicitly intensional notion of aboutness (Mai 2005, 599). These varying perspectives situate themselves along a distinction between inside and outside, and between being and doing. Some see the subject as an ontological entity distinguished by its differences from other entities: "This document is about economics as opposed to education." Others see the subject as a reference to an external context which determines the priority of a subject's facets: "Our users will want this document for its treatment of economic dimensions of educational policy."

Tensions Between Language and Discourse in North American Knowledge Organization

Ricoeur suggests that the prominence of semiotics in the twentieth century has prevented theorists of language from investigating discourse to its full potential. I would like to suggest that avoiding such a bias in knowledge organization gives us a richer view of North American research, which has already shown how specific acts of classification and description can be assembled and analyzed for patterns. Existing catalogue records have provided fruitful ground for emerging theories of the work (Smiraglia 2001), for the study of OCLC catalogue records (Miksa et al. 2006), and for hyperlinking patterns (Vaughan & Thelwall 2003). North American researchers are well-primed to assemble data on the discursive acts of knowledge organization, with the availability of OCLC, with Z39.50 access to numerous online catalogues, and the growing archives of harvested metadata records made available through the Online Archives Initiative. User tagging systems have provided rich new sources of user-centered knowledge organization in the areas of images and social bookmarking sites (Besiki & Jorgensen 2008; Kipp & Campbell 2006). And the Wayback Machine of the Internet Archive provides longitudinal evidence of knowledge organization on websites.

In particular, North American researchers could use the distinction between system and statement, between semiotics and semantics, as a way of analyzing a specific tension in knowledge organization between the local and universal: a tension that manifests itself on the one hand in the adaptation of specific needs to standard subject tools, and on the other in the rise of outsourcing in resource description.

Standard Tools and Special Interests

The field of Library and Information Science has traditionally excelled at the creation and study of standard tools which enable us to create uniform descriptions of subject content. Whether they be universal schemes such as Dewey's *Decimal Classification*, *The Library of Congress Classification* and the *Library of Congress Subject Headings,* or subject-specific schemes such as *MeSH*, the *NASA Thesaurus* and the *Art and Architecture Thesaurus*, these widely-used tools have large vocabularies and intricate and sophisticated syndetic structures that lend themselves to analysis as semiotic systems. This talent for building large subject systems has manifested itself on the Web in the form of library

application profiles for Dublin Core metadata, of the use of faceted classification in information architecture, and the growth of field-specific metadata schemes.

These impressive tools lend themselves to analysis as self-enclosed and consistent systems that, like Ricoeur's interpretation of *langue,* are the product of a specific discipline, in this case information science. We can assess their success at following principles of thesaurus construction, the consistency of their policies regarding equivalence, the rigour of their hierarchical structures, and their adherence to consistent facet orders. Furthermore, universal schemes such as DDC and LCSH have been analyzed for gender bias and other injustices (Olson 1998). Such studies typically use the structures, authorized vocabularies, and syndetic devices to chart the limits of what can or cannot be expressed within these subject languages, and showing how certain unarticulated presuppositions govern their use in the subject analysis process. These studies have had an important and beneficial impact on the revision and improvement of these schemes over the years: but while revising a vocabulary or classification may widen its range of possible utterances, it only addresses part of the issue. In addition to analyzing our potential utterances, we have to look at the utterances themselves.

Ricoeur's concept of semantics as a necessary partner to semiotics provides a conceptual frame for orienting numerous promising areas of research that look at what people actually do. First, focusing on acts helps us to investigate the enormous increase in personal information management that has extended from traditional practices onto the Web through Web 2.0 tools such as bookmarking systems. Equally important, the subject indexing and classification patterns of professional intermediaries can be explored in counterpoint to these new user-centered initiatives, through the availability of metadata harvesting sites.

North American researchers are also well-placed to address the problem of multilingual subject access. At the policy level, multilingual information access often appears to be a seamless process of translation at the system level, in which one term is mapped to an equivalent term in another language. In reality, the implementation of multilingual access is far more complex and ambivalent, and North American scholars are ideally placed to study how the uneven implementation of linguistic plurality policies is manifested in our knowledge organization systems. Researchers have noted unexpected anomalies in the transformations of bilingual catalogues (Arsenault & Menard 2007), and overt commitments to multicultural access are often belied by the rudimentary state of cataloguing and subject access to non-English documents.

Discourse and Outsourcing

North American researchers also need to address a growing trend of regarding information creation separately from its discursive context. The long-term contraction of funding for libraries has given rise to outsourcing, in which information communities end up suffering at the hands of their own skill in standards creation. The sharing of catalogue descriptions has been an intrinsic part of information organization at least since the rise of MARC, and the development of semantic Web tools also rests on the notion of data retooling and reuse. But cooperative cataloguing was never intended to suggest that library

records are solely the product of enclosed, self-referential descriptive processes, or that they can be created in a vacuum. Shared cataloguing ventures have always assumed that libraries would use the time saved by downloading basic records to shape those records into discursive artifacts appropriate for their libraries. Outsourcing, on the other hand, assumes that there is no discourse specific to the information context.

As our information environments grow progressively interlinked, North American researchers must continue to explore the available options and instruments that enable communities to assert their own needs and identities over collections described by universal standards. We need to track the degree to which options built into tools like DDC are actually employed; the extent to which Canadian libraries use the special areas of LCC set aside for Canadian history and literature. We need to theorize the optimal relations between human- and machine-readability in specific community settings.

The concerns for context, for diversity, and for flexibility have been present in the professional and academic communities of knowledge organization for a very long time. Likewise, the concerns for clear design, for international standards, and for enhancing universal access to information through technological and terminological continuities have figured large in North American information research. As the North American knowledge organization community gathers to assess its heritage and plan its future, Paul Ricoeur's theories of language, while far-removed from the particularities of our field, may provide a useful orienting distinction. Our concern for developing large, inclusive, interoperable and standard subject access tools must be counterbalanced by an equally close look at what information communities actually do with these tools. The conceptual clarity and discipline of information science that guide the creation of subject languages must co-exist alongside the inconsistent, haphazard and multidisciplinary context in which these languages are used.

References

Arsenault, Clement & Elaine Menard. 2007. Searching titles with initial articles in library analysis: a case study and search behaviour analysis. *Library resources & technical services* 51.3: 190-203.

Bates, Marcia J. 1999. The invisible substrate of information science. *Journal of the American Society for Information Science* 50(12): 1043-1050.

Beghtol, Clare. 1986. Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of documentation* 42.2: 84-113.

Besiki, Stvilia & Corinne Jorgensen. 2008. End-user collection building behavior in Flickr. *Proceedings of the American Society for Information Science and Technology*. 44: 1-20.

Blair, David C. 1990. *Language and representation in information retrieval*. Amsterdam: Elsevier Science.

Campbell, D. Grant. 2008. Derrida, logocentrism, and the concept of warrant on the Semantic Web. *Proceedings of the Tenth International ISKO Conference*. 11: 222-228.

Foskett, D.J. 1966. Introduction. *Documentation and the Organization of Knowledge*. by Jesse H. Shera. London: Crosby Lockwood & Son: xi-xxi.

Kipp, Margaret E.I. & D. Grant Campbell. 2006. Patterns and inconsistencies in collaborative tagging systems: an examination of tagging practices. *Proceedings of the ASIST Annual Meeting*.

Lancaster, F.W. 1986. *Vocabulary control for information retrieval*. 2nd ed. Arlington: Information Resources Press.

Mai, Jens-Erik. 1999. A postmodern theory of knowledge organization. *Proceedings of the ASIS Annual Meeting* 36: 547-56.

Mai, Jens-Erik. 2005. Analysis in indexing: document and domain centered approaches. *Information processing and management* 41: 599-611.

Maron, M.E. 1977. On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science*. 28: 38-43.

Miksa, Shawne, Moen, C., Polyakov, A. & Snyder, G. 2006. Metadata Assistance of the Functional Requirements for Bibliographic Records' Four User Tasks: a report on the MARC Content Designation Utilization (MCDU) Project. *Proceedings of the Ninth International ISKO Conference*. 10: 41-49.

Olson, Hope. 1998. Mapping beyond Dewey's boundaries: constructing classificatory space for marginalized knowledge domains. *Library trends* 47.2:233-254.

Ricoeur, Paul. 1976. *Interpretation theory: discourse and the surplus of meaning*. Fort Worth: Texas Christian University Press.

Saussure, Ferdinand de. 1985. The linguistic sign. *Semiotics: an introductory anthology*. Ed. Robert E. Innis. Bloomington: Indiana University Press: 28-46.

Smiraglia, Richard P. 2001. *The nature of 'a work': implications for the organization of knowledge*. Lanham: Scarecrow Press.

Svenonius, Elaine. 2000. *The intellectual foundation of information organization*. Cambridge: MIT Press.

Vaughan, Liwen & Mike Thelwall. 2003. Scholarly use of the Web: what are the key inducers of links to journal web sites? *Journal of the American Society for Information Science and Technology* 54.1: 29-38.