

Prosopography and the Organization of Literary Works: Toward a New Paradigm

Stephen Paling

Assistant Professor, School of Library and Information Studies, University of Wisconsin-Madison, 600 N. Park St., Madison, WI 53706-1403. E-mail: paling@wisc.edu

Introduction

Organizing literary works, particularly in terms of subject analysis, has traditionally presented the field with difficulties that are not necessarily present with other types of works. This paper is developmental in nature, and has three goals:

- Summarize and synthesize results from two separate but related studies that examined the organization of literary and nonliterary works from two different methodological perspectives.
- Describe a possible new paradigm suggested by the results of the studies.
- Discuss future research needed to elaborate and further test the proposed paradigm.

The studies in question also represent an attempt to avoid several assumptions:

- That users want access to literature at the level of the book.
- That existing cataloging tools are the best tools for organizing literary works.
- That quantitative and critical approaches need to be seen as antagonistic.

Rather than asking a more typical question such as, *What can we do with the tools we have?*, these studies are guided by the question, *What tools do we need to have in the first place?*

The Organization of Literary Works

The American Library Association has published *Guidelines on Subject Access to Individual Works of Fiction, Drama, etc.* (GSAFD) (ALA, 2000), but library catalogers often choose not to assign subject headings to literary works, and call numbers assigned to literary works often provide only broad groupings, e.g., by a broad range of dates. The ALA SAC Subcommittee on Fiction Guidelines (Bogan et al., 2003) issued a follow-up report on the GSAFD. The report referred to “the question of the legitimacy of the underpinnings of the [GSAFD],” and arguments “that novels don't have subjects and are not 'about' something, in the way of nonfiction works” (p. 5). The GSAFD and the subsequent report focused largely on fiction, and not on other types of literary works such as poetry. Other systems (e.g., Beghtol, 1994) have been proposed, but are not used in common practice. Databases of literary works exist, but do not enjoy widespread use.

The papers in this series build in part on a conceptualization begun in earlier work (Paling, 2004) that tied hermeneutics and deconstruction to the organization of information. Humanist scholars such as Derrida (1990) have made arguments similar to the arguments discussed by Bogan et al. (2003). Derrida suggested that all texts are inherently open in meaning, perhaps radically so, because they come to us free of their original context. Derrida argued, particularly of written texts, that the signifying form loses any absolute meaning,

by the possibility of its being repeated in the absence not only of its 'referent,' which is self-evident, but in the absence of a determined signified or the intention of actual signification, as well as of all intention of present communication. (p. 1175)

Gadamer (1976) took a similar position, arguing, “what is written is so distanced and detached from its author... that the grasping of the meaning of the text takes on something of the character of an independent productive act...” (p. 24). Openness of meaning could also create limits on our ability to describe the subject of texts. If readers and indexers indeed have trouble finding a clear or stable meaning in a text, then subject analysis for that text may be very difficult.

Taken together, the insights from Derrida and Gadamer provide us with an interesting point of departure. Their scholarship suggests that texts, rather than having a fixed and describable meaning, are far more mutable in meaning and only take on meaning in a specific historical context of reception. The insights from scholars such as Derrida and Gadamer, though, may not apply with equal force to all types of texts. The papers in this series are meant to begin exploring whether intellectual warrant exists to say that literary works are, indeed, different from nonliterary works, and whether a new paradigm for organizing literary works will be more effective than the current approach.

The Studies

The first study (Paling, 2011b) was a pilot study of literary community members that was meant in part to elicit data about the metadata that community members are likely to find useful in deciding whether or not to read a new literary work. The sample was relatively small (n=84), but appropriate for a pilot study. To make the discussion easier to read, this study will be referred to as *the metadata study*.

The metadata study tested five inductively built sets of potential metadata elements:

Bibliographic: Metadata elements that one would expect to find in standard bibliographic records (e.g., classification numbers and LCSH subject headings), as well as less-traditional uses of bibliographic records such as OCLC's FictionFinder project.

Faceted: Metadata elements found in, or based on, Beghtol's (1994) EFAS system. Included unusual metadata elements such as relationships between characters, e.g., mother-daughter, and the names of known fictional characters.

Popular: Metadata elements that one might find in popular sources such as online booksellers, including elements such as reader/user comments.

Shared: Metadata elements that did not belong clearly in any of the other sets because the elements are so common, e.g., the name of the author(-s), the length of the work, and the year in which a work was published.

Symbolic Capital: Metadata elements based on a work's place in literary writing

as a genre, e.g., the names of awards given to authors or works, or the names of authors published in the same literary magazine. This set of metadata elements was built on previous work done in this series of studies (Paling 2008; Paling 2009; Paling & Nilan 2006), which in turn built on work by Pierre Bourdieu (1992).

The element sets comprised a total of 42 elements, plus an Other category. An attempt was made to produce sets of elements that together would provide a broad and balanced selection for respondents to evaluate. However, the Other category was included in order to allow for unanticipated elements suggested by the respondents.

To keep the questions simple for the pilot study, the respondents were asked to answer Yes or No as to whether they would consider a particular potential metadata element when deciding whether or not to read a new literary work. *Literary work* was defined to include works that occur below the level of the books, e.g., individual poems and short stories.

Figure 1 shows a radial graph indicating the total percentage of Yes responses for each metadata element set. If each metadata element set had an equal percentage of Yes responses, then the pentagon would be symmetrical, which it clearly is not:

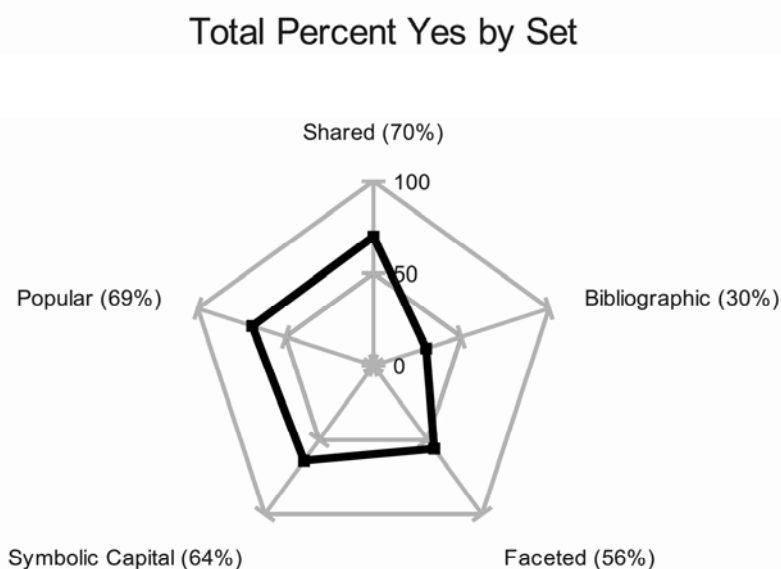


Figure 1: Total Percent Yes by Metadata Element Set.

The Bibliographic set was the only one to fall below 50% as a group. Two of the staples of library subject analysis, library-assigned subject headings and classification numbers, garnered 31% and 14% Yes responses respectively. Those results will be significant in light of the second study, which will be discussed later in this section.

The Symbolic Capital set was the only set of elements from the metadata study in which each potential element garnered Yes responses from a majority of the respondents. This is significant for at least two reasons. First, the Symbolic Capital set is based on the current line of research,

and the data provide early, positive indications of the usefulness of the elements in this set. Second, the Symbolic Capital set includes elements not typically found in bibliographic records or other sources of metadata. For example, the elements in the Symbolic Capital set include items such as a list of authors whose work has appeared in the same anthology as the work in question, and the title of a literary magazine in which the work has appeared. Paling (2011b) provides full details.

One thing that became clear was that not only do members of the literary community want access to literary works below the level of the book, they also want access to literary works across multiple instances. The evidence suggests that books, anthologies, magazines, etc., are simply *contingent collections*, i.e., collections put together for a particular rhetorical purpose, that may or may not themselves be useful as a unit of information. In many cases the needs of literary community members may match poorly with those contingent collections as community members seek individual, brief works such as poems or short stories.

The second study (Paling, 2011a) examined the concept of *indexability*, the ease with which a text can be organized through subject analysis for discovery or retrieval. This definition is complementary to Buckland's (1997) definition of *indexicality* as “the quality of having been placed in an organized, meaningful relationship with other evidence” (p. 806). The methodological approach taken in the study was strongly quantitative, but it was also heavily informed by work by scholars such as Derrida.

Fuzzy sets contrast with traditional, crisp sets by assigning members varying degrees of membership on a continuous scale from 0 (no membership) to 1 (full membership). The canonical example often used (Zadeh, 1965, p. 338) is the class of tall men. An extremely tall man, e.g., 7'2”, would have a very high membership value, while a man of below-average height, e.g., 5'0”, would have a much lower membership value. The idea examined in the second study is that different types of texts show different levels of membership in the fuzzy set of indexable texts. To make the discussion easier to read, this study will be referred to as *the fuzzy set study*. For full details see Paling (2011a).

The methodological approach in the fuzzy set study was very different than the approach taken in the metadata study. The approach was primarily quantitative, and involved the use of *visual analog scales*, VASs, to rate several properties of a set of purposively selected texts. The set included literary and non-literary texts, as well as one randomly generated text. VASs “employ a drawn or printed straight line of a specified length, with verbal anchors at each end, to represent a subjective state or stimulus... typically, a 100-mm line...” (Waltz, Strickland, and Lenz, 2005, p. 281). Participants mark the scale with a pen or pencil (although the scales can also be digital) to indicate the degree to which they experience a phenomenon.

The respondents, undergraduates at a large research university in the American Midwest, were randomly assigned to one of three groups. Respondents in all three groups were asked to provide terms that they thought described the subject of each text. Respondents in the first group was asked to rate the clarity of each text using a VAS whose anchors ranged from “Not At All Clear” to “Completely Clear.” Figure 2 shows a hypothetical response:



Figure 2: Hypothetical VAS Response.

The respondents in the second group were asked to use a similar VAS to indicate how confident they were about the subject descriptions they provided for the texts.

The respondents in the third group were asked to rank the texts based on clarity, but were not asked to use a VAS. Instead, they were asked to stack the texts in order of clarity, with the stacking serving to prevent ties. The simple ranking was used as a form of validation for the VASs.

The respondents tended to view the literary texts as less clear than the nonliterary texts, and expressed less confidence in their subject descriptions of the literary texts. The literary texts also garnered less consistent descriptions than did the nonliterary texts. Taken together, the results provide preliminary support for the idea that literary texts do indeed exhibit lower indexability than non-literary texts. In other words, they would have lower membership value in the fuzzy set of indexable texts.

Taken together, the two studies present an interesting set of preliminary conclusions. The literary community members interviewed for the metadata study show very limited desire for the staples of library subject analysis, subject headings and classification numbers. Similarly, the undergraduates in the fuzzy set study found the subjects of literary works less clear than the subjects of nonliterary works, and expressed less confidence in their own attempts to describe the subjects of literary works. These preliminary results suggest that subject analysis of literary works may not be the most effective tool for organizing those works, and that any paradigm that emphasizes books as units of information may fail to meet the needs of people seeking literary works.

Prosopography as a New Paradigm

If the traditional paradigm that calls for the (at least occasional) assignment of subject headings or broad classifications to literary works does not meet the needs of users, we are still left with the question of what paradigm might meet those needs more effectively. This paper posits for further exploration the *prosopographic paradigm*. Stone (1971) defined prosopography as “the investigation of the common background characteristics of a group of actors in history by means of a collective study of their lives” (p. 46). The data used include “kinship, socio-economic position, residence, occupation, etc.” (Carney, 1973, p. 156). Keats-Rohan, Borymchuk, and Fernholz (n.d.) emphasized that, “Central to prosopography are the concepts of individuals and groups - or networks” (Key Ideas page), and observed that it is based on structured biographical information (What is Prosopography page). The definition proposed here, though, would be slightly broader, and would comprise *networks of people, works, and publications*.

A relatively simple example from the work of poet Russell Edson can serve as an illustration. His poem “Clouds” appeared in *The Very Thing That Happens* (Edson, 1964), published by New Directions, as well as in *The Tunnel: Selected Poems of Russell Edson* (Edson, 1994), published

by Oberlin College Press. Denise Levertov wrote an introduction to *The Very Thing That Happens*. New Directions also published work by William Carlos Williams and H.D. (Hilda Doolittle). So this network, or perhaps a fragment of a larger network, would contain at least the following elements:

- Russell Edson.
- *The Very Thing That Happens*.
- *The Tunnel: Selected Poems of Russell Edson*.
- “Clouds”.
- Denise Levertov.
- William Carlos Williams.
- H.D.
- New Directions.
- Oberlin College Press.

The network could also be viewed ontologically as comprising the following relationships (and others):

- publishedBy.
- publishedIn.
- writtenBy.

The network would also comprise the following contingent collections (and others):

- Works by Russell Edson.
- Authors published by New Directions.
- Works with writing by Denise Levertov.
- Anthologies in which the poem “Clouds” has appeared.
- Literary magazines in which the poem “Clouds” has appeared.

The last two potential metadata elements are hypothetical in the case of “Clouds”, but are nonetheless supported by the data from the metadata study. In that study, 75% of the respondents indicated that they would consider the name of a literary magazine in which a work had been published when deciding whether to read that work, and 67% indicated the same about the list of authors whose work had also appeared in an anthology with the work. Compare that to the 14% and 31% garnered, respectively, by classification numbers and library-assigned subject headings. This also fits with the results of the fuzzy sets study, which points to potential difficulties with the subject analysis of literary works.

The prosopographic paradigm suggested here is at once both simple and rich. It is simple in the sense that the description of these relationships does not require a great deal of interpretation. Information about where authors have published, for example, is relatively easy to obtain. Even information that may be more difficult to obtain, for example, whether a poem has appeared in an anthology, is nonetheless often unambiguous. But at the same time, such a network can contain rich ontological connections that will be useful to many users. Some of these potential

metadata elements are easily used in traditional bibliographic tools, but others are not. For example, the fact that a poem in a volume of poetry was previously published in a particular literary magazine, and later in an anthology, is rarely included in standard bibliographic records. The preliminary evidence suggests, then, that standard bibliographic records provide only a partial match with the needs of users looking for literature.

It is worth noting, though, that other metadata points of interest to the respondents in the metadata study may prove to be a good deal less practical. For example, 82% of the respondents in the metadata study indicated interest in abstracts or similar summaries of literary works. For books, that might be feasible. But providing abstracts or summaries at the level of individual, brief works such as poems would present clear problems. Many authors would be reluctant to provide summaries for work they consider finely crafted, and writing summaries for a large number of brief works would probably be more than most cataloging departments or metadata projects could reasonably provide. Similarly, a determination of the era or time of day described in a work, although of interest to 72% of the respondents in the metadata study, may be difficult at best, and, in some cases, deliberately ambiguous for effect.

There is a sense in which this is conceptually simple. While users may express interest in a wide variety of metadata elements, we have a professional imperative to also consider what is actually feasible to deliver. Part of what the evidence from the two studies suggests is that, while we have done a reasonable job working with what is feasible, we have paid too little attention to what is desired by our users. For example, it has proven feasible, by and large, to provide authoritative versions of authors' names. The author's name was the single most interesting potential metadata element to the respondents in the metadata study, garnering 93% Yes responses. But other potential metadata elements, such as the publication history of a poem or short story, while obtainable and of interest to the respondents in the study, are not regularly present in standard bibliographic records.

Future Research

It is important to remember that the conclusions in the two studies discussed in this paper are preliminary, and that additional research will be necessary to validate and refine those results.

Conceptually, it is easier to start with potential follow-ups to the metadata study. Working at the productive conjunction of the desirable and the feasible would require us to pursue at least several refinements to the metadata study:

- Pursuing evidence of whether members of the literary community have different desires or needs for metadata than do non-members such as readers of popular fiction.
- Determining not simply whether respondents are potentially interested in each potential metadata element, but also the amount of importance they attach to each element.
- What metadata elements will be feasible to provide for users.

Refinements to the fuzzy set study will be more complicated, but equally important for our understanding of how human beings view different types of texts. The goal in this line of research is to begin an agenda of basic research into human perception of different types of texts, and what those perceptions can tell us about the amenability of different types of texts to subject

analysis. The fuzzy set study provided very preliminary evidence that people do, indeed, view some texts as potentially more indexable than others. In other words, some texts have higher membership value in the set of indexable texts.

In addition to gathering additional data to take this line of research beyond the pilot stage, several other refinements are possible. The most obvious refinement would be to compare the perceptions of end users with the perceptions of professional indexers and catalogers. Any difference between those perceptions would raise additional research questions. For example, if catalogers and indexers perceive clear meaning in a literary text where end users do not, we would have to ask how we might go about reconciling those two sets of perceptions. That would involve difficult questions such as whether the subject perceptions of catalogers and indexers are actually useful for users seeking literary works.

Another set of potentially productive refinements could come from utilizing different models for fuzzy sets. As an example, we can return to the set of tall men. Within some margin of error, we can know a man's height, 5'7", 6'2", etc. The mathematical fuzziness comes in our interpretation of the membership value each height provides in the set of tall men. But with texts we probably face a somewhat different situation. Human perceptions of the clarity of a text, or users' expressions of their own confidence in describing a text's subject matter, will probably vary more than we would expect to see with a simpler measure such as human height. Put somewhat differently, we need a model for fuzzy sets that allow us to take into account a greater range of human judgment than would likely be present with other measures.

Fate mi (2011) described a model for stochastic intuitionist fuzzy sets (Sifts). He contrasted this model with the original model proposed by Zadeh (1965):

Entropy measures in the stochastic systems are the degree of uncertainty which result from randomness, while in the theory of fuzzy sets uncertainty is due to the vagueness and complexity of the system. The the fuzzy entropy can be considered as the degree of uncertainty caused by personal judgment. (p. 748)

Fate mi also proposed that in order to fully describe an intuitionist fuzzy set, we need to have two of the following three functions: the membership function, the non-membership function, and the hesitation margin (p. 749). The features of the Sifts model that allows the modeling of human judgment and hesitation bring rich possibilities for describing human perception of literary and nonliterary texts. The simpler model used in the fuzzy set study established the potential utility of fuzzy sets, but future studies need to take into account subsequent advancements in fuzzy set theory such as the advancements made by Fate mi. Those newer models may allow us to more effectively model the uncertainty expressed by end users about literary texts.

Another productive research question will address whether the literary/nonliterary axis is the only axis along which we can construct fuzzy sets. There could, for example, be an expert/non-expert axis, as well. The preliminary results discussed here suggest that nonliterary texts, i.e., non-fiction prose, are more indexable than literary texts. But the fuzzy set study did not include in the purposive sample of texts highly specialized prose that one might encounter in academic

fields such as engineering or physics. There are at least several possibilities to explore. Do non-experts see those technical texts as having clear subjects, even if the non-experts don't understand the subjects? For example, a non-expert with little or no knowledge of electrostatic forces might nonetheless perceive the text as being clear for someone who does understand the subject, and the non-expert may feel confident in naming electrostatic forces as the subject.

To summarize, the two studies discussed here provide evidence for the following very preliminary conclusions:

- End users perceive literary texts as being less clear than nonliterary texts.
- End users are less confident in their ability to describe the subject of literary texts than nonliterary texts.
- Traditional bibliographic records are only a partial match for user needs in seeking literary works.
- Traditional staples of library subject analysis hold weak appeal for members of the literary community.

But these preliminary results need refinement. This paper represents an attempt to begin establishing a research agenda that will make those refinements possible, and that can lead to a possible new paradigm for organizing literary works.

References

ALA. (2000). *Guidelines on subject access to individual works of fiction, drama, etc.* (2nd ed.). Chicago, IL: American Library Association.

Beghtol, C. (1994). *The classification of fiction: The development of a system based on theoretical principles*. Methuen, NJ: Scarecrow.

Bogan, R., Desi, S.M., Gable, L., Gilchrist, J., Earn, S., Hoffman, S., Loins, D., Miller, J., Rob are, L., Walters, C. and Sewerage, M. (2003). *Final report of SAC Subcommittee on Fiction Guidelines*. Chicago, IL: American Library Association.

Bourdieu, Pierre. 1992. *The rules of art: genesis and structure of the literary field*. Stanford, CA: Stanford University.

Auckland, M.K. (1997). What is a 'document'?. *Journal of the American Society for Information Science*, 48, 804-809.

Edson, R. (1964). *The very thing that happens: Fables and drawings*. New York: New Directions.

Edson, R. (1994). *The tunnel: Selected poems of Russell Edson*. Oberlin, OH: Oberlin College Press.

Derrida, J. (1990). Signature event context. In P. Bizzes and B. Herzegovina (Eds.), *The Rhetorical Tradition* (pp. 1168-1184). Boston, MA: Bedford Books.

Fate mi, A. (2011). Entropy of stochastic intuitionist fuzzy sets. *Journal of Applied Science*, 11, 748-751.

Gadamer, H.G. (1976). On the scope and function of hermeneutics reflection. In D.E. Liege (ed. and trans.), *Philosophical Hermeneutics* (pp. 18-44). Berkeley, CA: University of California.

Paling, S. (2004). Classification, rhetoric, and the classificatory horizon. *Library Trends*, 52, 588-603.

Paling, S. (2008). Technology, genres, and value change: literary authors and aesthetic use of information technology. *Journal of the American Society for Information Science and Technology*, 59, 1238-1251.

Paling, S. (2009). Mapping techno-literary spaces: Adapting multiple correspondence analysis for literature and art informatics. *Information Research*, 14(2). Retrieved February 1, 2010, from <http://informational/Ir/14-2/paper401.HTML>.

Paling, S. (2011a). Fuzzy sets and the organization of literary texts: A conceptualization and two-stage pilot study. *Journal of Documentation*, 67(4), 601-623.

Paling, S. (2011b). Developing a metadata element set for organizing literary works: a survey of the American literary community. *Knowledge Organization*, 38(3), 262-277.

Paling, S., Nilan, M. (2006). Technology, values, and genre change: the case of little magazines. *Journal of the American Society for Information Science and Technology* 57, 862-872.

Waltz, C.F., Strickland, O.L. And Lenz, E.R. (2005). *Measurement in nursing and health research*, (3rd ed.). New York: Springer.

Zadeh, L.A. (1965). Fuzzy sets. *Information and control*, 8, 338-353.