**Lilium Rajan**

**What and Where is Ambiguity in Categorization?**

**Abstract**

Categorization systems organize the world along various lines for specific purposes. This sorting inevitably creates ambiguities when the structure can’t neatly address particular realities. While we necessarily accept these bits of confusion or misidentification as the fallout of our work, we lack clear models for pointing to where and when these cases of misfit occur. Discussing this matter is made more challenging by the terminologies used by practitioners and end users. This model, tentatively called the chain of surrogation, attempts to map the actors, objects and processes of categorization work and assign working definitions to each stage of the categorization process. Doing may allow us to understand what it means for something to be ambiguous in a categorization system and who, or what, is responsible. Then we may ask how we might resolve or retain that ambiguity.

**1 Introduction**

Categorization systems organize the world along various lines for various purposes. Classification is the most formal categorization practice, employing standards and rules to guide professional members of a discipline in assigning descriptions. Less formal categorization practices require similar transformations of observations of things into descriptions of those things but often rely on simpler tools for guiding the process.

My work focuses on ambiguity in categorization systems broadly, asking what and where it might be. I follow the work of critical knowledge organization scholars who have given attention to both pragmatic and ethical issues related to categorization in information systems. Much of their work has considered the impacts of classification on marginalized groups of people. It has tended to focus on the power of naming and issues of exclusion, tracing the wide gap between community approaches to self-representation and classificatory terminology or relationship structures. In exploring ambiguity more broadly, I hope to offer an approach that allows us to ask similar questions about a wider range of groups and items, both human and non-human. This approach also allows us to ask questions about cases where certainty may not be possible or desirable.

To anchor my research questions and aid in my analysis, I have developed a four-stage view of categorization work informed by library and archives practices, which I tentatively call a *chain of surrogation*. This model allows me to point to different kinds and sources of ambiguity and ask questions about who or what might be responsible for this kind of uncertainty. This model also attempts to disambiguate many terms that are used in knowledge organization, including *identity, representation, description, object,* and *resource.* The definitions I propose here are not meant to suggest that any of these terms should our could have only one meaning. Instead, they allow my work to proceed with consistency and clarity. As this model evolves, changes in the terminology may be warranted.

**Diagram of Chain of Surrogation. Model moves from item to description, both concrete, through identity and representation, both abstract. Three primary kinds of people interact with this system, starting with system designers and ending with users. The identity stage includes a tension between approaches: identicality or best fit. The representation stage attempts to map the identity to the priorities of the system, which can include considerations of both item access and resource allocation. 


**

**2 Chain of Surrogation**

This model attempts to address how we sort in systems that are designed by institutions (sometimes called “organizations”, a term we will exclude in this model). Institutions generally do this work to provide access to the things that are sorted and/or to allocate resources related to those things. The goals, values and requirements of the system determine how things are represented. The infrastructures of a system, including standards, rules, forms and software, determine how things are described.

Categorization work starts with an item and ends with the description of the item that is recorded for later access. In the chain of surrogation, we suppose that initial *item* is examined to determine its *identity.* The identity must be matched to the *representation* goals of the system before one can arrive at a *description.* This process, which is usually performed by a professional, traverses the material and the conceptual. Ambiguity may be seen at each of these stages of categorization, as well in the end-user experience of encountering descriptions or the items themselves. Depending on where the ambiguity occurs or is experienced, there may be different approaches to its resolution—or, perhaps, preservation.

**2.1 Item (concrete)**

The thing that needs to be accounted for in a system, which may or may not be held by the organization that is doing the accounting. This may be a physical or non-physical item (a book vs. a digital file, for example). In systems that categorize people, a person or community may be understood at this level. *Synonyms*: Resource, person, individual, exemplar, entity, object, document. Note: it is unclear to me at this stage whether “data” may also be a synonym here.

**2.2 Identity (conceptual)**

What the Item is in relation to other items. Dual understandings of “identity” make this move from item to identity challenging. Furner (2009) points out that can see identity as either *that which is identical,* that is, exactly the item; OR we may see identity as *belonging to a group based on shared characteristics*. We understand both approaches as valid and try to reconcile them by highlighting both what is different about an item in relation to other items, and to declare how it is most similar to other items. In either case, declaring identity is attempting fidelity to the item.

Identity of an item may be drawn from the item itself in whole or in part (a book, for example, has many identifying characteristics; a person can state their identity, which still may consist of both unique and matching conceptions of identity). In other cases, the identity must be determined by the person doing the description and categorization work. Determining the identity and the appropriate means of describing the item requires observation of the item through the lens of the system.

**2.3 Representation (conceptual)**

Representations are created by an institution and are constrained by system designs.

Representation is the goal of a categorization system. It is the structure for accounting, whatever the goals and values of the system may be. Representation is an attempt to match items to some resources, be they space, money, or energy. Representation facilitates access *to* the items or *for* the items accounted for by the system. Access is deemed successful by a variety of measures including efficiency, interoperability, accuracy, stability, accessibility, warrant and cohesiveness/synthesis.

**2.4 Description/Categorization (concrete)**

Descriptions (or determinations of category) are created by practitioners and encountered by users. The means by which an item is represented in a system is via its description. A description attempts to render an identity for an item in accordance with the representation goals of a given system. The description is the end result of the work that a person (or in automated systems a machine, though I am not considering these cases) does after considering the item’s identity in relation to the representation goals and rules of the system.

There are cases where description is arrived at via direct engagement with a system, as when a person self-represents on a form or in a census. It is not clear how this kind of direct entry is distinct from professional entry, though it seems reasonable to expect that professionals have a wider range of goals and values, as well as detailed understandings of the rules that determine their description work.

**3 Ambiguity: Thing or Experience?**

There are two ways of considering ambiguity. First, we can see ambiguity as a thing, or as a property of a thing. This assumption is embedded in the thing-oriented discipline of bibliographic classification. In this field, we look to resources to derive our descriptions. We try to remain objective and descriptive in our use of concepts. We want to be able to create rules that can be used by multiple professionals and be consistently understood by most or all users. There is inevitable compromise in this, but it is generally acceptable and domain specific (Hjørlund, 2009). However, human identities are not a domain. People are not resources in which one can find precise attributes that allow us to make the right call. Furner grappled with this in his analysis of identity (2009).

This tension suggests that in some cases, ambiguity can be understood as a thing or a characteristic of thing, but more often it is an experience. We encounter, as professionals or end users, and wonder to ourselves what precisely something is. In some cases, we can point to the sources of our confusion, but the confusion is ours. The world is not entirely concrete, yet knowledge organization wishes to render it so. What are we failing to capture? Gorichanaz (2016) has characterized this split as one between *veritas,* which emphasizes factuality, and *alethia,* or uncovering, and suggests that knowledge organization is preoccupied with veritas at the expense of the much more challenging alethia. This may map well to the differences between ambiguity as thing versus experience.

I am interested in capturing both possible versions of ambiguity in my research, as they may suggest distinct approaches. This model allows us to capture ambiguity as a property or an experience at various stages of the descriptive process.

**4 Where is Ambiguity?**

My research is primarily concerned with how and where ambiguity is found or experienced in categorization systems. While my initial work has focused on classification and categorization systems in library and archives settings, these questions may apply as well to categorization in other information systems and to issues around the construction and uses of data.

The simplest way of thinking about ambiguity in these systems is to ask: what doesn’t fit? We might think of a category of *miscellaneous* or *other* and ask what is held there, and why? What are the characteristics of these odds and ends? However, most information systems do not offer much insight here. Either the rules of the systems do not allow for such chaos or the practitioners responsible for sorting are disinclined to leave such ragged edges. The following framings of ambiguity as an end-user issue; a practitioner issue; a system design issue; or an item issue, allow us to ask distinct questions. The chain of surrogation suggests a flow from the concrete item through the system description, with the end user as the final point. I have arranged these framings from the most extrinsic questions of item identity, held in the end user, to the most intrinsic questions about the item itself, each framing connecting to the set of questions prior.

I do not intend this flow to suggest that ambiguity in fact always lies in the item itself, but to offer a path that starts with the person who is assumed to have the least familiarity with the goals and limitations of the system.

**4.1 Ambiguity as an end-user issue**

Given the absence of a simple junk drawer to mine, must ask our question slightly differently: what strikes a viewer as peculiar, odd or confusing upon encountering a description? What is difficult or impossible to find, even if it is somewhere in the system? Perhaps we might say that ambiguity is a user problem, arising when someone doesn’t know how to make the right query, or doesn’t understand the representation goals and limitations of the system. Even so, we would want to look at the descriptions encountered by the end user to see where the problem might lie. Is it in the words used to describe an item? Is it in their arrangement amongst other items?

**4.2 Ambiguity as a practitioner issue**

Imagine that we discover a description that has been deemed ambiguous or challenging to an end user. Where might the problem have arisen, other than the user’s own limitations? The person who has done the description may be consulted: what approach did they take when describing this item? What resources were they highlighting? What was their concept of identity for the item, and from what was it drawn? What rules and standards governed their approach? These questions anchor the issue of ambiguity in the tension between observing the item and translating it into a description.

**4.3 Ambiguity as a system design issue**

The person who created the description probably did not create all these governing rules—they simply put them to work. How then might we consider the culpability of the representation system itself in creating or forcing ambiguity (or, more provocatively, allowing it)? The design of the tools used to facilitate representation might also be considered: Why are they created the way they are? What priorities are they surfacing? What are they unable to address? What are they not interested in addressing?

**4.4 Ambiguity as an item issue**

While the previous sites of ambiguity are probably the most prevalent, it is also worth asking what role the item itself has in its ambiguous descriptive rendering. In cases where self-identification is allowed, where humans are permitted to enter their identity into the system, ambiguity can be an act of resistance when the system is not trusted, or an act of confusion when the purposes of the system are not well understood. Items may also be said to be ambiguous in relation to a system when they fail to provide adequate information that is legible to the system (fragments) or when their uses are beyond the understanding of those who describe them. In this case, one might fault the practitioner for suffering from uncertainty. It is not clear to what depth or degree an item can be considered constitutively ambiguous outside of the system of representation. My work allows for this possibility without taking a firm stance on whether an item can be inherently ambiguous.

**5 Conclusion**

This model, still a work in progress, attempts to provide an overview of the process of categorization in multiple kinds of information systems. This process model may be useful to a variety of researchers as an aid in pointing to particular moments and transformations. It was developed more particularly to allow me to think and write more clearly about the very blurry concept of ambiguity in relation to things that are categorized. It is not clear to me whether this model has extended utility, though I hope that if taken up by others it could provide a means of anchoring conversations about related but disparate problems in categorization. Providing a clear yet robust visual model of the chain of surrogation remains a challenge. The version provided does not attempt to include the full range of inputs, outputs and synonyms that have informed this work, and as such may be viewed as either obvious or too vague. I welcome feedback on how to render this model and make it useful to other research areas.

**References**

Furner, J. (2009). Interrogating ‘Identity’: A Philosophical Approach to an Enduring Issue in Knowledge Organization. *Knowledge Organization* 36: 3–16.

Gorichanaz, T. (2017). Information and experience, a dialogue. *Journal of Documentation* 73(3): 500-508.

Hjørland, B. (2009). Concept Theory. *Journal of the American Society for Information Science and Technology* 60(8): 1519-536.

Jacob, E. K. (2004). Classification and categorization: a difference that makes a difference. *Library Trends*, *52*(3), 515.



Olson, H.A. (2002). *The Power to Name: Locating the Limits of Subject Representation in Libraries*. Boston, Ma.: Kluwer Academic Publishers.

Rajan, L. (2017). Ambiguity in Knowledge Organization: Four Proposed Types. *NASKO* 6: 239-247.

Rajan, L. (2018). Historical ambiguity: a lens for approaching outdated terms. *Challenges and Opportunities for Knowledge Organization in the Digital Age:* 256-264. Ergon-Verlag.

Star, S.L., and Bowker, G. (2007). Enacting silence: Residual categories as a challenge for ethics, information systems, and communication. *Ethics and Information Technology* 9: 273–280. <https://doi.org/10.1007/s10676-007-9141-7>