

Lilium Rajan
University of Washington Information School

Ambiguity in Knowledge Organization: Four Proposed Types

Abstract

Classification and categorization order by creating or seeking certainty. Yet inevitably we encounter things that defy ready placement, which we may label *other* or *miscellaneous*, or force into another category. The literature of knowledge organization recognizes the consequences of classification and misrepresentation, but has not systematically outlined what circumstances or conditions render a thing ambiguous to those who would seek to describe it.

This paper proposes four major sources or type of ambiguity in classification. While examples of these types may be found in many disciplines and settings, they have in common similar requirements for accurate or improved representation. *Multiplicity* is a source of ambiguity when a resource or object requires more terms to describe than the system allows. *Emergence* is ambiguity that arises when phenomena, from medical observation to literary genre, is at an early stage of description and thus unstable. *Privacy-related* ambiguity is that which stems from a gap of understanding or trust between those classifying and what is being classified, particularly in human communities. *Conditional* ambiguity arises when something requires narrative due to conditional contexts such as temporality or geography. This term also describes things that have dichotomous or fragmentary identities that are not easily represented by most systems.

These types of ambiguity may arise in formal and informal organization systems. While observing these types of ambiguity may not offer immediate or feasible solutions, it may allow us to discuss their unique challenges and to better understand their manifestations across disciplines.

Introduction

Classification and other forms of categorization are employed to help us make sense of the world. We sort according to our needs and enforce our decisions through systems of varying formality. Recent scholarship has interrogated many systems and their categories, bringing to light diverse consequences of classification. Systems of all sorts require compromises, of which designers and users may not be aware.

Compromises become particularly visible in their representations of uncertainty, as when resources are relegated to the uncomfortable category of “other.” Most of the time, however, uncertainty remains invisible as we seek to disambiguate or at least render our systems to appear as though we have. Knowledge organization and classification seek to eliminate ambiguity for good reasons, but if we don't know how to think deeply about our sources of confusion and where they lie, we will be continuously misclassifying. We accept compromises, but is it enough to declare that we are making compromises without knowing which ones we are making?

Extensive scholarship in the past two decades has illuminated the many ways in which identities are othered or marginalized in classification systems. Most of this research has focused on representations of human identity groups. In this paper I will introduce a conceptual scheme of ambiguity that may allow us to better see what is happening in various situations where we are challenged by the task of representing or classifying. I outline four major types of ambiguity, which are illustrated by many

different forms in which ambiguity can arise. These types are tentatively described as Multiplicity, Emergence, Privacy and Conditional Ambiguity. My goal in creating categories of ambiguity is to begin documenting examples of these, to test their use and their limitations. Eventually, I hope we may find our way toward solutions for better representations, recognizing that ambiguity is to some extent in the eye of the beholder and that competing ideals of representation may be forever at odds. Nonetheless, being able to consider the issues from the perspective of multiple stakeholders may help knowledge organization practitioners anticipate and alleviate troubles associated with ambiguity.

Method

This paper presents a vocabulary that I have developed in the course of my research into the nature or contents of otherness. While there is abundant evidence and nuanced discussion of the consequences of othering, especially as related to representations of human identity, there is little discussion in the literature of bibliographic classification about non-human entities that are othered. I initially framed othering as a consequence of confusion on the part of those who classify, which led me to ask: what is so confusing? This approach led me to reframe the question in terms of ambiguity, a state that can lead to confusion and othering. As I encountered examples of others across formal and informal systems in the literature and in encounters in daily life, I noticed characteristics in common between some of them, which have coalesced into the four types of ambiguity proposed here. Issues of power and intentional control as related to othering can be seen within some of these types of ambiguity. However, these issues are not the primary focus of this work. Instead, types of ambiguity are arranged in relation to their primary causes and potential solutions.

Classification beyond the library

This paper will address aspects of ambiguity that relate to many forms of representation. While most of my work has been centered in LIS, I have encountered relevant examples of ambiguity in other settings as well. Rather than addressing a context such as bibliographic classification, I am seeking to describe the aspects of ambiguity that arise across contexts of representation, whether in formal or informal systems. Because the concept of ambiguity holds in many settings, in this paper I will sometimes refer to the objects of classification or categorization as *things* rather than as resources.

Although much of the research in the field explores representation in formal classification systems, I will address these questions as related to categorization, following Jacob (2004). This view of sorting considers relationships between things but is less concerned with hierarchies and the systems in which the categories are used. A more general approach may make this analysis useful to both formal and informal categorization settings, ranging from detailed classification systems to situations of routine but unorganized categorization such as the use of jargon in a professional setting. However, the formal requirements of classification may introduce another layer of tension in addressing ambiguity.

The other subjects

All of our sorting activities assume or seek certainty. In classification, there should be a place for everything. What then of those things that inevitably defy ready categorization? There are many ways that these challenges are handled, from being scuttled into a junk drawer category such as “other” or “miscellaneous,” or being forced into another category that may or may not have a clear relationship to the resource in question. They may even be ignored or discarded entirely. In addition to being confusing or useless, these misrepresentations can perpetuate and enforce power over marginalized identities, as has been illustrated by Olson (2002), Higgins (2016) and Fox (2016), among others.

There are three primary ways that things end up as others. In the first case, one might know what something is, but the system simply has no place for it. For example the first instance of a new media type arrives in a collection, and the instrument of classification has not yet addressed the proper handling of this form. While the person classifying has no trouble describing what the item is, none of the available terms fits.

In another case, the thing itself might be entirely unknown or indescribable by the classifier, perhaps due to missing context. Archives and museums that have fragments or under-documented objects may contain others of this sort.

Finally, a thing might be considered *other* due to ambiguity, which is the primary focus of my research. In order to understand how ambiguous others are handled by classification systems, it is helpful to consider the many ways in which a thing can be ambiguous.

In this paper, I will explore the variety of contexts in which *otherness* manifests, in hopes that a high-level understanding of the factors that lead to othering may offer solutions for all stakeholders in the work of representation. Otherness has often been seen in struggles around terms for such aspects of human identity as race, gender, sexuality, ethnicity or disability. Rather than focusing on how to properly represent these categories, this paper will discuss othering in terms of ambiguity or non-categorization, which may be particularly relevant in cases of intersectionality.

Stakeholders in representation: knowledge organizers, users and resources

Questions of representation are of interest to three primary groups. First are those who author and implement classification systems, who seek to represent their world through a set of terms and to assign those terms consistently to their body of resources. Next are users who must refer to the system to do their work and who look to the system to understand the contents of their domain. Finally, these questions of representation are of interest to the resources (things) themselves, when the resources are related to human identity, as well illustrated in Adler’s work on representations of transgender terms (2009), Furner’s work on racial classifications in the DDC (2009) and Fox’s ontogeny of intersex (2016).

These tensions can also arise in situations where the resource is authored or created by a person or community who has an interest in how the object is represented. Research in indigenous knowledge, for example, has addressed representations of community knowledge in library and museum contexts (Duarte and Belarde-Lewis 2015). Ambiguity or otherness, then, may be a matter of perspective. The system author may see ambiguity or an other in a resource she is unable to readily describe.

The user may find ambiguity in an absence of the thing sought, in a lack of representation. The resource may experience ambiguity in othering or other forms of misrepresentation where the assigned identity is inaccurate, sometimes intentionally.

Otherness: treatment or condition?

We tend to see otherness as a treatment (and an inadequate one at that) rather than as a condition. That is, we understand that otherness is a matter of perspective and usually stems from an inhospitable method of description rather than a fundamentally confused or confusing thing. Ambiguity, on the other hand, is more generally unsettling, an experience of questioning (“what is this?”) rather than declaration (“this is other”). By considering ambiguity broadly rather than otherness specifically, we will address challenges to representation that stem from conditions that confuse us as we go about representing. This approach will may allow us to consider our treatments to avoid othering. Ambiguity can be seen or experienced in situations that have little to do with human identity as well. This paper will attempt to map out major sources of ambiguity, agnostic of theme. While this approach will offer few specific solutions, it may allow us to address ambiguity systematically or to see solutions in cases with similar types of ambiguity. This exploration is by no means complete and will benefit from discussion with practitioners from the field of knowledge organization and beyond.

The case for ambiguity: beyond the other

Examining ambiguity allows us to see patterns in categorization that range from the accidental, to the confused, to the intentionally limiting. This line of questioning is an extension of the questions brought forth by Star and Bowker in their discussion of residuality (2007). This paper addressed the existence of things that are unclassifiable due to issues of temporality and personal experience or perspective, such as fluctuating physical pain. It is not possible to pin these things down in representation. One wouldn’t want to. However, this impossibility also renders these realities as formally non-existent. More recent work by Feinberg, Carter and Bullard has explored the residual through redesign of the metadata of video collections (2014). In addition to drawing on the concept of residuality, they consider Anzaldúa’s mestiza consciousness, which addresses identities at crossroads (1987). These concepts not only illuminate the problems of indeterminate identities but also may offer approaches for working with them.

Current treatment of ambiguous things

There are several ways that ambiguity is handled in systems. An ambiguous resource may be up-posted, or described as something that it is a part of, in which case its particulars are lost. In other cases, it may be falsely named or given a term that is deemed synonymous, in which case it may disappear, while counting as a representation of something else entirely. In some cases, it may be declared *other* or some variant, usually nested within a classification. Less formal representation, such as arises in conversation, can sometimes approach ambiguity with a sort of shrug or avoidance, or with a longer narrative description.

Consequences of ambiguity in representation

Ambiguity and *others* pose a challenge to the illusion of complete order that we seek in classification. While we may reasonably hope to find more accurate ways of representing knowledge, it is also important to understand that many solutions are compromises that have consequences. Othering can marginalize subjects, but it can also suggest ways of evaluating the consequences of classification. When is it good to up-post, and when might it be worth retaining the unsettling label of “other”? What is happening when a classifier engages in false synonymy, and which areas are particularly subject to this kind of misunderstanding? When might we best serve a community or a resource by using less explicit terms, by retaining a certain amount of ambiguity without falsely naming?

It can be helpful to highlight some of the primary consequences of ambiguity. As is true of all categorization and representation, systems and their contents are often looked to as expressions of reality. Therefore, even misrepresentations can be viewed as real, especially when there is no way for a user to understand the limitations of the categorization (Bowker and Star, 1999). The most common concern about othering is disappearance or invisibility through misrepresentation or underrepresentation. In cases where ambiguous resources are up-posted, there may be resultant over-representation of a superior class.

All forms of representation can be linked to resource allocation, which is one of the reasons that we seek fidelity in our representations. Whether we are trying to make decisions about how to develop collections or are counting who works in a company, knowing what exists allows us to decide what we need. Ambiguous representations will inevitably shift our perceptions of what is possible and what is necessary.

The concept of organization is implicitly a move toward efficiency. Ambiguity makes a system less efficient, but handling ambiguity is also disruptive to efficiency. It requires time and resources, yet the specific time and resources required to address or resolve ambiguity can be different depending on context. This analysis of ambiguity does not eliminate the need for resources in solving these issues but may pave the way to smoother resolutions.

While there are challenges associated with ambiguity in systems, in some circumstances retaining ambiguity may offer us some possibilities. Retaining some form of other may preserve subclasses that may prove relevant and useful. Additionally, in instances where ambiguity is an intentional aspect of a resource, retaining ambiguity in representation may be in line with the community needs.

Organizing Ambiguity: Forms and Types

Ambiguity can be difficult to discuss because it is often experienced as a sort of confusion. It can arise from different aspects of the resource in question, ranging from issues of size (where the very small fragment is as confusing as the immense collection) to issues of perspective or power. In this paper, we begin by identifying the forms or characteristics of resources that can contribute to ambiguity. These forms can then be organized into four primary types of ambiguity that are encountered in categorization and classification: Multiplicity, Emergence, Privacy, and Conditional. These four primary types and their embedded forms are characterized by their potential solutions—or lack thereof.

Multiplicity

Multiplicity describes resources that have so many relevant categories or subjects that they cannot be adequately represented by most systems. Most ambiguity of this sort is seen in things that are either *hybrid* or *immense*. Hybridity is a combination of two or more things, retaining defining characteristics of the constituent parts, and can describe everything from racial identity to literary genre. A thing that is ambiguous due to its immensity is too vast and heterogenous to readily classify, as in a large set of mixed data.

While hybridity may be easier to describe than immensity, they can both be difficult to represent in a system that allows only limited terms for representation. In an ideal setting, however, they could be accurately conceptually represented through a complete list of terms, however inefficient that might be. We recognize that an exhaustive inventory of terms would introduce many new challenges to comprehension and retrieval and that representation must take these factors into account. Nonetheless, it can be useful to consider these forms of ambiguity as a type with a hypothetical solution.

Emergence

Emergence describes those things that are ambiguous due to their novelty but which may become more fully understood as members of a new category over time. The *nascent* phenomenon or resource is considered to be too new to be useful to the system. Moving out of this ambiguous state usually requires evidence of repetition of that form over time, as seen in the concept of literary or cultural warrant (Beghtol 1986). In science contexts, we may require replication of observations or some form of peer review to push an emergent identity out of ambiguous limbo. Emergence is frequently seen in literary and music genres, such as steampunk literature. It can also be seen in the invention of new forms of digital media. For example, how might we characterize short videos that are designed to disappear after viewing?

Ambiguity due to emergence is usually resolved with the passage of time and thus requires patience and attentive observation. In the meantime, resources may be misrepresented or ignored and may be lost to future use even if their more appropriate representations introduced into the system at a later date.

Privacy

Privacy refers to ambiguity that arises when the motivations of a system are in some way at odds with the needs of the represented (Nissenbaum 2004). In the case of human identities, this may be due to a lack of trust in the system, as when identification of race or sexuality is elicited. It can also be seen in museum setting when objects associated with community knowledge are removed from their original contexts. This kind of ambiguity can be seen as either *invisible* or *defiant*. Invisibility occurs when direct representation has been difficult or impossible and a community has come to use coded language or different terminology. One might sense that the terms have another meaning, such as when references to spinsters or unmarried aunts are used as hints of homosexuality. As is seen in this example, these associations can be internal or external euphemisms. Defiant ambiguity occurs when a thing or phenomenon intentionally

avoids classification, generally because the community from which it emerges wants to avoid limiting it or making it publicly discoverable.

There are at least two solutions in cases of privacy-related ambiguity. First, the administrators of system might seek to increase the trust of the communities represented in order to gain access to less ambiguous representations. This may be as simple as allowing community to submit or use their choice of terms of representation (as seen in Facebook's gender options) or may require more relationship building. In other cases, the system, such as a museum, may preserve the public ambiguity of the resource in accordance with the wishes of the source community. An interesting approach to issues of ambiguity and trust can be seen in the work anthropologist Aaron Glass (2015) has done with Kwakwaka'wakw communities around cultural objects that are housed in museums around the world. Rather than create single representations of cultural objects that may be associated with private or community knowledge, Glass and his partners created two metadata schemas dealing with provenance and allowed the communities to decide whether the indigenous provenance metadata would be shared with the museums or the public.

Conditional

Conditional ambiguity describes the many forms of confusion that arise from context. It may be seen as the "not otherwise specified" of types of ambiguity, as it is associated with the widest range of forms. It broadly describes forms of ambiguity in which representation must grapple with the conditions in which the resource is described. In some cases ambiguity arises because the resource is a *fragment*, a portion of something larger, where it is clear that significant information is missing and it is unclear what role the fragment plays (or even what it is a part of). Related to fragmentary ambiguity is *corrupt* ambiguity, in which a resource may be understood but is in a condition that prevents it from being grouped with non-corrupt members of a like class. Accurate representation will want to account for both the incompleteness, or damage, of the resources and any possible conjecture involved in their description.

Opaque ambiguity arises when the person doing the representation cannot adequately describe the thing in question, though it may be complete, due to inaccessibility. This may be due to issues of translation or cultural context. Representation of these resources may be something akin to a best guess. *Historical* ambiguity may arise when a resource employs a term that does not have current standing. For example, the term Asperger's has been eliminated from the DSM-V as a classification related to the autism spectrum, yet the term is still widely used by people in the autism community. We can also see the consequences of changes in use of terms in a classification system, such as eugenics, as issues of historical ambiguity (Tennis 2012). *Misfit* ambiguity is seen when something has been located with other items that seem to be unlike. It may be unclear whether the resource was previously misclassified or whether the classifier is missing the original context for the arrangement. Examples of misfits include a biography intentionally shelved in fiction or a book of prose in the poetry section. This could also include ethnic or racial classifications, such as the designation of Indians as Caucasians in some systems.

Dichotomous ambiguity is a subset of multiplicity, whether hybrid or immense, in which some of the terms required for representation are considered impossibly

simultaneous in the logic of the system and thus require further explanation. For example, one might encounter a text that is both fiction and non-fiction, or a person who identifies as both female and male. While a system may allow these terms to coexist, they require description of relationship to make sense.

Finally, conditional ambiguity can arise when an accurate representation of a thing is dependent on *temporal* or *geographical* conditions. Bowker and Star's residually of pain can be seen as an example of this type of ambiguity. Artifacts may also exhibit this kind of ambiguity, as when an object is used for different purposes depending on location and time. For example, a wine bottle may become a rolling pin in a kitchen, a vase in a bedroom, and a water scoop by a stream.

Cases of conditional ambiguity require some form of narrative in order to be accurately represented, which is at odds with the economical requirements of classification. Where expanded narrative or explanatory metadata is not possible, it might be possible to arrive at a workable compromise of representation through negotiation with stakeholders, but this will most likely still produce a compromised representation that satisfies only some requirements or stakeholders.

Conclusion

Precision and certainty seek economy for many reasons: clarity, interoperability, and size constraints. But truly ambiguous resources or entities challenge the efficiency constraints of a system by requiring something closer to a natural-language description to accurately represent. The goals of categorization tend toward efficiency, which is at odds with the representation needs of resources beset with multiplicity. The demands of efficiency are even more challenging in cases of conditional ambiguity.

Practitioners of knowledge organization are likely familiar with many or all of these sources of ambiguity and have found a variety of solutions to them as they are encountered. However, we are often motivated by a desire to disambiguate, to provide access through representation. In this desire we may overlook the qualities that challenge us, assuming that we already know the compromises that must be made. My hope in introducing this overview of ambiguity is that we can begin to work toward more accurate representations by engaging the very qualities or aspects. We may find new approaches from across systems and disciplines as we communicate our solutions and compromises to these types of ambiguity. We may also find ourselves advocating for retention of some aspects of ambiguity in the form of privacy shielding representations or narratives.

References

- Adler, Melissa. 2009. "Transcending Library Catalogs: A Comparative Study of Controlled Terms in Library of Congress Subject Headings and User-Generated Tags in LibraryThing for Transgender Books." *Journal of Web Librarianship* 3: 309–331. <https://doi.org/10.1080/19322900903341099>
- Anzaldúa, Gloria. 1987. *Borderlands: the new mestiza = La frontera*. San Francisco, Ca.: Spinsters/Aunt Lute.
- Beghtol, Clare. 1986. "Semantic validity: concepts of warrant in bibliographic classification systems." *Library Resources and Technical Services* 30: 109-125.

- Bowker, Geoffrey, and Susan Leigh Star. 1999. *Sorting Things Out: Classification and Its Consequences*. Cambridge, Ma: MIT Press.
- Duarte, Marisa Elena, and Miranda Belarde-Lewis. 2015. "Imagining: Creating Spaces for Indigenous Ontologies." *Cataloging & Classification Quarterly* 53:677-702.
- Feinberg, Melanie, Daniel Carter, and Julia Bullard. 2014. "A story without end: writing the residual into descriptive infrastructure." *DIS 2014 - Proceedings of the 2014 ACM SIGCHI Conference on Designing Interactive Systems, Vancouver, Canada, June 21-15, 2014*. 385–394. <https://doi.org/10.1145/2598510.2598553>
- Fox, Melodie. 2016. "Subjects in Doubt: The Ontogeny of Intersex in the Dewey Decimal Classification." *Knowledge Organization* 43: 581-593.
- Furner, Jonathan. 2009. "Interrogating 'Identity': A Philosophical Approach to an Enduring Issue in Knowledge Organization." *Knowledge Organization* 36: 3–16.
- Glass, Aaron. 2015. "Indigenous Ontologies, Digital Futures: Plural Provenances and the Kwakwaka'wakw collection in Berlin and Beyond." In *Museum as Process: Translating Local and Global Knowledges*, ed. Raymond Silverman, 19-44. London: Routledge.
- Higgins, Molly. 2016. "Totally Invisible: Asian American Representation in the Dewey Decimal Classification, 1876-1996." *Knowledge Organization* 43: 609-621.
- Jacob, Elin K. 2004. "Classification and Categorization: A Difference that Makes a Difference." *Library Trends* 52: 515-540.
- Nissenbaum, Helen. 2004. "Privacy as contextual integrity." *Washington Law Review* 79: 119–157.
- Olson, Hope A. 2002. *The Power to Name: Locating the Limits of Subject Representation in Libraries*. Boston, Ma.: Kluwer Academic Publishers.
- Star, Susan Leigh, and Geoffrey Bowker. 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>
- Tennis, Joseph T. 2012. "The Strange Case of Eugenics: A Subject's Ontogeny in a Long-lived Classification Scheme and the Question of Collocative Integrity." *Journal of the American Society for Information Science and Technology* 63: 1350-359.