A Theoretical Framework for Operationalizing Basic Level Categories in Knowledge Organization Research

Abstract
Research on categories indicates that superordinate categories lack informativeness because they are represented by only a few attributes while subordinate categories lack cognitive economy because they are represented by too many attributes (e.g., Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Basic level categories balance informativeness and cognitive economy: They represent the most attributes common to category members and the fewest attributes shared across categories. Green (2006) has suggested that the universality of basic level categories can be used for building crosswalks between classificatory systems. However, studies of basic level categories in KO systems have assumed that the notion of a basic level category is understood and have failed to operationalize the notion of "basic level category" before applying it in the analysis of user-generated vocabularies. Heidegger's (1953/1996) notion of handiness (i.e., zuhandenheit, or being "at hand") can provide a framework for understanding the unstable and relational nature of basic level categories and for operationalizing basic level categories in KO research.

A substantial body of research indicates that basic level categories are the first categories formed during perception of the environment, the first learned by children, and those most used in language (e.g., Mervis & Crisafi, 1982; Tversky & Hemenway, 1983). In his seminal article, Roger Brown (1958), who is often identified as the first researcher to have theorized the existence of a basic level of categorization (e.g., Lakoff, 1986; Murphy 2002), described how parents teach their children to name and categorize objects. He argued that the "name of a thing, the one that tells what it 'really' is, is the name that constitutes the referent as it needs to be constituted for most purposes" (p. 17). Brown points out that, while some names given to an object are "re-categorizations" that represent "acts of imagination" (p. 17), the names of things as they really are serve as true representations of the referent. Brown's contention that category names of things represent things "as they are" -- as the thing "needs to be constituted for most purposes" -- foreshadowed Eleanor Rosch's theory of basic level categories.

Drawing inspiration from Brown's (1958) observations and from Wittgenstein's (1953/1963) discussions of language use, Rosch and her colleagues (Rosch & Mervis, 1975; Rosch et al., 1976) developed the theory of basic level categories. Rosch (1978) argued that, in the perceived world, information-rich bundles of perceptual and functional attributes form natural discontinuities and that basic cuts in categorization are made at these discontinuities (p. 31). Working from these assumptions, Rosch et al. (1976; Rosch, 1978) defined basic level categories as most inclusive because they convey "the most information, possess the highest cue validity, and are, thus, the most differentiated from one another" (Rosch et al., 1976, p. 383). According to Rosch and Mervis (1975):

Basic objects (for example, chair, car) are the most inclusive level of abstraction at which categories can mirror the correlational structure … of the environment and the most inclusive level at which there can be many attributes common to all or most members of the categories. (p. 586)

Thus, while experimental studies have suggested that superordinate categories lack informativeness because they are represented by only a few attributes and that subordinate categories lack cognitive economy because they are represented by too many attributes (Rosch & Mervis, 1975), basic level categories provide a balance between informativeness and cognitive economy because they represent the most inclusive level of abstraction: the level at which the greatest number of attributes are common to category members (e.g.,
and the fewest number of attributes are shared across categories (e.g., dog and cat). In the years since Rosch introduced the notion of basic level categories, her theory has been applied in a wide variety of cognitive and linguistic studies, including word use and free naming (e.g., Tanaka & Taylor, 1991; Markman & Wisniewsky, 1997; Blewitt, 1983); American Sign Language for the deaf (e.g., Newport & Bellugi, 1978); categorization of images (e.g., Murphy & Smith, 1982); and environmental scenes (e.g., Tversky & Hemenway, 1983; 1984).

Studies of basic level categories have indicated that entry level categorization (i.e., the level of abstraction that is perceived or identified first during the individual's interactions with the environment) may be modulated by two factors: the individual's domain-specific knowledge and the typicality of an exemplar for a corresponding category. Thus Jolicoeur et al. (1984) claim that the "level at which objects are identified first depends on typicality" (p. 271), while Rips et al. (2006) speculate that factors responsible for an individual's preference for a basic level term "may include not only those associated with its basic-level category but also the larger set of background causes that govern the individual's environment" (p. 9). It has frequently been argued that both the influence of an individual's experience and the typicality of the object within a given environment may actually shift the entry level of categorization to the subordinate level (e.g., Rosch et al., 1976; Jolicoeur et al., 1984; Tanaka & Taylor, 1991; Belke et al., 2010). As Brown (1958) observed:

It seems likely that things are first named so as to categorize them in a maximally useful way. For most purposes Referent A is a spoon rather than a piece of silverware, and Referent B a dime rather than a metal object. The same referent may have its most useful categorization on one level (Prince) for one group (the family) and on another level (dog) for another group (strangers). The categorization that is most useful for very young children (money) may change as they grow older (dime and nickel). (p. 20; emphasis in original)

Within the framework of the experientialist view, conceptual structure is meaningful because it is embodied, that is, it arises from, and is tied to, our perceptual bodily experiences" (Lakoff, 1987, p. 267). For Lakoff, embodiment is coupled with "collective biological capacities and … physical and social experiences [of individuals] as beings functioning in … [their] environment" (p. 267). The experientialist view assumes the existence of an external structure in "our bodily experience" (p. 267), which arises from the structured "preconceptual experiences" (p. 267) of individuals. Lakoff (see also Lakoff & Johnson, 1980) suggests that there are at least two kinds of structure that are implicated in preconceptual experiences: basic level structure and kinesthetic image-schematic structure. Basic level structures are comprised of categories that represent the convergence of gestalt perception, the capacity for bodily movement, and the development of rich mental images, while kinesthetic image-schematic structures are the external structures that constantly appear in everyday bodily experiences -- structures such as paths, containers, or front-back and part-whole relationships (p. 267). Lakoff also theorizes that, through the use of metaphors and the projection of basic level categories to superordinate and subordinate categories, individuals extend these basic level and image-schematic structures to create abstract conceptual structures. However, the experientialist approach does not address the relational and unstable nature of preconceptual structures such as basic level categories: For example, the perception and understanding of categories and category boundaries can differ substantially across individuals, situations, and cultures and even change dramatically over time, and the experientialist view cannot account for such variability and instability.

LIS research has tended to confine research on basic level categories to two areas of investigation: How the basic level of categories is represented in knowledge organization systems (e.g., Iyer, 1995; Fernandez & Eastman, 1990; Green, 2006); and how basic level terms are used to represent the conceptual content of resources such as images (e.g.,
Jorgensen, 2003; Rorissa & Iyer, 2008; Rorissa, 2008; Yoon, 2009; Schmidt & Stock, 2009). Green, Bean and Hudon (2002) investigated the occurrence of basic level categories in four different systems of knowledge organization: a bilingual thesaurus (Canadian Literacy Thesaurus), a biomedical vocabulary (Unified Medical Language System), and two ontologies (ThoughtTreasure and WordNet). Analysis of a random sample of terms taken from these resources, which included the semantic scope of a term as well as its placement in the original hierarchy, demonstrated that equivalence across systems occurred significantly at the basic level rather than at subordinate or superordinate levels. Interestingly, basic level terms were observed most often in the Canadian Literacy Thesaurus and least often in the Unified Medical Language System. Analysis using goodness-of-fit tests confirmed that, in these four systems, basic level terms were more likely to have equivalent terms across systems than were subordinate or superordinate terms.

The most extensive study of basic level categories reported in the LIS literature was conducted by Green (2006). She identified a set of basic level terms by analyzing the structure of WordNet\footnote{http://wordnet.princeton.edu}, a lexical database of English words in which 117,000 groups of synonyms (or \textit{synsets}) are linked by hierarchical (i.e., superordinate-subordinate) or part-whole relationships. Based on her analysis, Green (2006, p. 8) developed six general criteria for the identification of basic level categories:

- Length and structure of lexical units (i.e., whether the lexical unit is a simple word, a compound word, or a phrase);
- Level of occurrence within the WordNet tree structure;
- Number of links to parts;
- Frequency of usage;
- Total number of links to other categories; and
- Number of immediate children (i.e., subordinate categories one level down) and overall number of children (all subordinates).

Applying these criteria to an analysis of every leaf node in WordNet's network of 59,692 noun synsets, Green identified 7,168 basic level categories.

To investigate the hypothesis that category terms at the basic level were more universal and therefore more likely to co-occur across classificatory systems than either subordinate or superordinate category terms, Green investigated the occurrence of basic level categories in a sample of 28 thesauri in ten subject domains (i.e., agriculture, education, engineering, environment, graphic materials, health, information science, legislation, political science, population science, and water sciences). She randomly selected ten category terms from each thesaurus and expanded these terms using the relational structure of the thesaurus from which each term was drawn. For each category term in these hierarchical expansions, the closest corresponding category term was identified in each of the other thesauri for that subject domain to produce category pairs consisting of one category term from the hierarchical expansion in one thesaurus and the closest corresponding term in a second thesaurus. Green then analyzed all category pairs to determine if the two category terms were equivalent and applied the criteria for basic level categories generated in the analysis of WordNet to identify basic level category terms.

Green's findings appear to support the hypothesis that basic level terms are universal. She suggests that the practical outcome of her study is the direction it provides for building crosswalks between classificatory systems: Given that basic level categories are significantly more likely to have exact equivalents across systems than categories not at the
basic level, developing mappings across classificatory systems should emphasize basic level categories to generate the “cleanest mappings” (2006, p. 12). However, even though the apparent universality of basic level categories in systems of knowledge organization and representation would seem to indicate their universality in indexing as well, basic level categories may not be appropriate for supporting collocation precisely because they have such a high probability of occurrence.

Applying the experientialist approach, Tennis (2005) proposed a user-centered classificatory structure that could potentially improve the static nature of traditional classification systems through incorporation of user-generated terms that are linked to the concepts defined in classification system. However, Tennis’s proposal inherits the problems of the experientialist approach, including, in particular, the unstable and relational nature of an individual’s preconceptual structure. Moreover, the instability and variability of individual preconceptual structures may also affect the process of classification and the application of principles of classification that govern the structure and relationship of classes (Jacob, 2004).

Although the universality of basic level terms is generally accepted by researchers in LIS, many of whom have suggested the positive outcomes of considering basic level terms in information representation and organization (Iyer, 1995; Bates, 1998; Hoenkamp, Stegeman & Schomaker, 1999), few studies have actually tested the utility of basic level terms in information organization and retrieval. Furthermore, studies that have investigated basic level terms generally have not provided an account of how the notion of basic level categories was actually operationalized (e.g., Golder & Huberman; Munk & Mork, 2007). Several studies (e.g., Rorissa & Iyer, 2008; Rorissa, 2008; Schmidt & Stock, 2009) have adapted the coding scheme that Green devised in her analysis of WordNet (Green et al., 2002; Green, 2006). However, even though WordNet has become the “de facto standard because it is freely available for research” (Ide & Wilks, 2007, p. 52; emphasis in original), the database itself was constructed based on paradigmatic relationships between nouns and verbs, which can lead to difficulties when attempting to determine the exact semantic level of a particular term. Thus, for example, because WordNet does not include adjectives, Schmidt and Stock (2009) were forced to identify the noun form to which an adjective was related in order to identify hierarchical relationships (e.g., to identify happy with the hierarchical position of happiness).

Problems associated with Green’s (2006) operationalization of basic level terms based on the hierarchical structure of synsets in WordNet are not limited to the determination of a term’s exact semantic level. More importantly, this approach masks the importance of the context in which a category term occurs and thus fails to take into account the relational nature -- the “handiness” or Zuhandenheit (Heidegger, 1953/1996, p. 65) -- of basic level terms.

Because the relational nature of basic level categories is either masked or ignored when the research focus is on the category as a term, Heidegger’s notion of handiness offers critical insights for more effective operationalization of basic level categories. Heidegger defines "handiness" -- the notion of being “at hand” or "ready-to-hand" -- as the “ontological categorical definition of beings as they are ,in themselves” (p. 67; emphasis in original). Handiness is the pre-reflective understanding of a thing as useful, or "ready-to-hand", on the basis of what is objectively present" (p. 67) -- on the basis of a thing as it is and how it is or, as Brown (1958) observed, "how it needs to be constituted for most purposes" (p. 17). In Heidegger’s account, handiness is the subject of assertion which ultimately plays some role in practical inference” (Rorty, 1991, p. 32) and thus reflects and contributes to the individual’s understanding of the world as it is. Thus handiness is an
assertion that is grounded in inference based on the realization of things as they are. In sharp contrast to the Platonic account, which limits assertion to theoretical inferences, Heidegger's approach reflects its relationship to American pragmatism by extending assertion to the individual's everyday activities within the physical world.

For Heidegger, a useful thing is ʻessentially ʻsomething in order to...ʻ... such as serviceability, helpfulness, usability, handiness, [which] constitute a totality of useful things” (p. 64). He argues that the meaning of ʻin order to” includes a ʻreference of something to something” (p. 64; emphasis in original). In particular, things are seen as useful ʻin terms of their belonging to other useful things” (p. 64):

... writing materials, pen, ink, paper, desk blotter, table, lamp, furniture, windows, doors, room[:] These „things” never show themselves initially by themselves, in order then to fill out a room as a sum of real things. What we encounter as nearest to us, although we do not grasp it thematically, is the room, not as what is „between the four walls” in a geometrical, spatial sense, but rather as material for living. (p. 64)

Thus the notion of handiness cannot be grasped or explained theoretically; rather, it can only be apprehended as an association that „makes use of things” and „guides our operations and gives them their specific thingly quality” (p. 65). For example, a hammer is associated with hammering; but the act of hammering itself „discovers the specific „handiness’ of the hammer” (p. 65). For Heidegger, then, the handiness of a useful thing is discovered in its association with the „work to be produced” (p. 65):

The shoe to be produced is for wearing (footgear), the clock is made for telling time. The work which we primarily encounter when we deal with things and take care of them—what we are at work with—always already lets us encounter the what-for of its usability which essentially belongs to it. (p. 65; emphasis in original)

Within the framework of Heidegger's notion of handiness, basic level categories are „useful things” that „reveal” themselves as “things” by their handiness. Moreover, it is possible that it is exactly this „thingly” (p. 65) quality that underlies the predominance of basic level categories in everyday vocabulary. While the predominance of basic level categories may be due, in part, to a general understanding of „work to be produced” (p. 65), Heidegger sees the work produced not only as a reference to the „what-for of its usability” of a thing and the „whereof of which it consists” (p. 66), but also to the „surrounding world of nature” (p. 67). He provides as an example the use of a clock, arguing that:

When we look at the clock, we tacitly use the „position of the sun” according to which the official astronomical regulation of time is carried out. The surrounding world of nature is also at hand in the usage of clock equipment which is at first inconspicuously at hand. (p. 67)

For Heidegger, the handiness of things in the „work world nearest to us” (p. 67) is based on the function of discovery and depends upon the „way we are absorbed[:] innerworldly beings that are brought along together with their constitutive references are discoverable in varying degrees of explicitness and with a varying [degree of] attentive penetration” (p. 67). In other words, the use of „things” is directly related to the individual's perception and understanding of those things and, as such, those things do not have an independent meaning that exists outside the individual's understanding. As Dewey argued (1997), the dependence on human perception and understanding to ascribe meaning to a thing represents the knowledge that enables individuals to „adapt [the] environment to our needs...[and] aims and desires to the situation in which we live” (p. 217). Thus, for Dewey, knowledge is comprised not only of what „we are now conscious of” but also of the „dispositions we consciously use in understanding what now happens” (p. 217).

Within the framework of Heidegger's notion of handiness, variation is both natural and phenomenological in that perception and understanding -- and thus the meaning of "things" --arise out of the individual's contextualized experiences of engaging with objects. Basic
level categories vary across individuals and across cultures because of differences in the everyday experiences and activities of individuals. The implications of handiness for basic level category research in LIS is obvious: The operationalization of basic level categories based on the hierarchical structure of normalized vocabularies cannot account for variations in the relational structures of individuals, much less variations in shared conceptual structures across cultures. A more effective approach to operationalizing basic level categories must take into account the contextualized experiences of individuals actively engaging with objects in order to assess their understanding of an object's "thingly quality" (p. 65) and its relationship to other objects in the everyday world of work.

References