Eastman, C. (1994). Anthropological Perspectives on Classification Systems. 5th ASIS SIG/CR Classification Research Workshop, 69-78. doi:10.7152/acro.v5i1.13777

Anthropological Perspectives on Classification Systems

Caroline M. Eastman
Department of Computer Science
University of South Carolina
Columbia, South Carolina 29208
eastman@cs.scarolina.edu

Robin M. Carter
4165 East Buchanan Drive
Columbia, South Carolina 29206

Some anthropological perspectives on classification systems are presented here. Such systems have been extensively studied by cognitive anthropologists; the area of work is often referred to as ethnoscience.We summarize work related to three specific domains: color terms, biological classifications, and kinship terminology. Some generalizations that can be drawn from this work involve categories, features, and the structure of classification systems; variations in terminology and system metrics are also discussed.The conclusions and generalizations from this work are relevant to the design and use of classification systems to support information retrieval applications.

INTRODUCTION

Classification systems are used in all human cultures. Some support ordinary discourse and everyday activities; these systems usually do not have explicit formal structure and may not exist in written form. Others support the organization and retrieval of information from manual or computerized systems. These generally exist in written form and have a formal structure.However, both informal and formal classification systems have common characteristics and elements.Examination of informal systems can provide guidance for the development and use of formal systems for retrieval purposes.

There has been extensive study of human classification systems by anthropologists and other social scientists; they have examined a variety of specific classification systems and also general principles and characteristics. The anthropological study of human knowledge systems, including classification systems, is usually referred to as ethnoscience; it is part of the anthropological subfield of cognitive anthropology.

Ethnoscientific studies have examined a variety of knowledge domains across a large number of human cultures.We consider here three specific domains in detail: color terms, ethnobiological classification, and kinship terminology. These three domains are relevant to all human cultures and have been extensively studied.We first summarize what is known about classification systems for these three domains. We then discuss some general characteristics concerning classification systems and how people use them.

COLOR

The domain of color is one for which all humans confront the same physical reality, in which colors can be represented as wavelengths. People with normal eyesight receive the same physical stimuli. However, we do not all describe the colors we perceive in the same way. Some cultures use a wider

ISSN: 2324-9773

Eastman, C. (1994). Anthropological Perspectives on Classification Systems. 5th ASIS SIG/CR Classification Research Workshop, 69-78. doi:10.7152/acro.v5i1.13777

variety of color terms than others, and there are differences in how colors are classified.However, there are definite patterns and common elements observed in color systems.

Colors can be represented as a grid showing a variation of wavelengths (hues) and brightness. Each color term represents a region on this grid containing a focal point which is generally agreed to be described by that color term. Although there is general agreement on the foci both across cultures and within cultures, there is much less agreement on the boundaries.Because the boundaries are unclear, fuzzy set models have been used by some to describe color terms (Kay and McDaniel, 1978).

The classic reference on color terms is Berlin and Kay (1969), which presents a thorough cross cultural comparison of color terms.They focus on basic color terms; these can be characterized by several properties. Monolexemic (simple) names are used. The color is not included in another term; this would exclude *turquoise* as a special case of *blue*.The term applies to all objects rather than a special set of objects; this would exclude *blond* in English.The term is known and psychologically salient to all speakers. Eleven basic color terms have been identified: *white, black, red, green, yellow, blue, brown, purple, pink, orange,* and *gray.* These appear to be universal perceptual categories.

Some cultures use only a subset of these terms rather than all of them. These subsets are not arbitrary; only a small number (22) of the possible subsets (2,048) of these color terms were observed in the different cultures examined by Berlin and Kay. Not using all of these terms does not mean that speakers of that language do not perceive the same differences among colors that others do, merely that they classify them differently.

Berlin and Kay propose a developmental sequence for languages which describes the observed subsets of basic color terms used.

Stage I: black, white
Stage II: black, white, red
Stage III: black, white, red, either green or yellow
Stage IV: black, white, red, green, yellow
Stage V: black, white, red, green, yellow, blue
Stage VI: black, white, red, green, yellow, blue, brown
Stage VII: black, white, red, green, yellow, blue, brown, at least one of purple, pink, orange, and gray

Berlin and Kay present data for 98 different languages.There is at least one language at each of these stages.There is some evidence that these stages represent developmental stages which occur as languages change and develop. Kay and McDaniel (1978) reexamine and extend this general framework of analysis. They argue that color terms result from the biological nature of human visual perception. Our visual receptors are most sensitive to *red, yellow, blue,* and *green* colors. These four colors, along with *black* and *white*, should thus occupy a dominant position in our color systems.These are the colors used in the first five stages. Other colors, including the remaining basic color terms, can be regarded as fuzzy combinations of these colors.

ISSN: 2324-9773

Eastman, C. (1994). Anthropological Perspectives on Classification Systems. 5th ASIS SIG/CR Classification Research Workshop, 69-78. doi:10.7152/acro.v5i1.13777

BIOLOGICAL CLASSIFICATION

An ethnobiological classification is a classification system for organisms used in a particular culture. It is less formal and less exhaustive than a scientific classification system. Berlin (1992) presents a comprehensive view of current knowledge about ethnobiological classification systems. He presents and discusses several generalizations about the structure of such systems. These are summarized below.

All cultures name and classify some of the organisms found in their local environment. Not all organisms are included; those which are included are those which are most useful and/or noticeable. These categorizations are based upon observable morphological and behavioral characteristics. Ethnobiological classifications are organized as hierarchies containing four to six levels.

The levels or ranks correspond roughly to those found in scientific classification systems; they are kingdom, life-form, intermediate, generic, specific, and varietal. The correspondence is not exact. For example, *seagull* is a generic folk taxon which includes several scientific genera. The most numerous taxa in classification systems are those of generic rank; most of these are monotypic and are not further subdivided.

Consider American terms for *birds*. Most Americans recognize such birds as *robins, eagles, sparrows, crows, seagulls, chickens* and *penguins*. This classification can be compared to scientific classification. Some of these terms correspond to single species (robin), some to a single genus (crow), and others to a group of genera (seagull). Some are readily observed (robin); others are part of the folk taxonomy even for people who may have never actually seen one in the wild (penguin). Birds are part of a broader group, *animals*. Ornithologists and birders recognize more different kinds of birds and use more elaborate classifications for them.

Ethnobiological classifications do not always form clean and elegant hierarchies.They may lack a root. Some taxa are covert; they appear to be recognized but are not named. The same name may be used at more than one level. For example, Americans sometimes use *animal* to refer to *mammals* only and sometimes use it in a broader sense.The hierarchy may be a tangled hierarchy. For example, most *dogwoods* are *trees*, but some are not. *Chickens* are not only *birds* but also *domestic animals*.

KINSHIP TERMINOLOGY

All humans share the same biology; genetics and reproduction are the same in all cultures. We all have relatives. However, the ways in which we classify and interact with relatives differ from culture to culture. This is reflected in the terms we use for relatives. Kinship terms do not always translate exactly from one language to another. Schusky (1983) presents a concise summary of kinship terminology and analysis.

Analysis of kinship terminology often starts with a set of basic kin terms which can be combined in various ways. The set of terms commonly used is mother (Mo), father (Fa), brother (Br), sister (Si), son (So), daughter (Da), husband (Hu), and wife (Wi). For example, mother's mother is written MoMo. Analysis is performed from the perspective of a particular individual, referred to

ISSN: 2324-9773

Eastman, C. (1994). Anthropological Perspectives on Classification Systems. 5th ASIS SIG/CR Classification Research Workshop, 69-78. doi:10.7152/acro.v5i1.13777

as *ego*. Kinship terms often represent sets of basic combinations. For example, the English term *grandmother* includes both MoMo and FaMo. The number of combinations for *cousin* is much larger. Analysis of a kinship system involves first determining what possible combinations a particular term refers to and then determining how this set can be characterized. For example, *grandmother* is female, a direct ancestor, and two generations removed from ego.These three properties represent the three dimensions relevant to analysis of the American kinship system: sex, generation, and lineality.

Most cultures recognize both *affinal* and *consanguineal* ties. Affinal relationships involve marital ties, such as *wife* and *sister-in-law.* Consanguineal ties involve blood relationships, such as *father* and *granddaughter.* In addition most societies provide mechanisms to create additional ties, such as adoption.

Groupings of kin which are used and appear natural in one culture may differ from those used elsewhere. Americans distinguish *lineal* and *collateral* consanguineal relatives. Lineal relatives are direct ancestors or descendants, such as *daughter.* Collateral relatives are other consanguineal relatives, such as *aunt.*This distinction is not made in many cultures. For example, a father may be grouped with his brothers. There are also distinctions made by others which appear unnatural and unnecessary to many Americans, such as the concept of *cross cousin.* A cross cousin is the child of a father's sister or of a mother's brother.

Even kinship terms can be fuzzy or indeterminate in some cases. For example, the boundaries of the American term *cousin* are not clearly defined. Cousins are not restricted as to generation or degree of lineality, and people differ in how far from themselves they are willing to use the term cousin. Terms can vary. One person might use the term *second cousin* and another might use *cousin once removed* to refer to the same relationship. Lakoff (1987) describes in detail the complex usage of the term *mother* in American society. This term is used in a wide variety of contexts, including *foster mother, biological mother, surrogate mother*, and *mother earth.*

FEATURES

Anthropologists examining classification systems often attempt to isolate and identify features of the terms contained within them.This process is often referred to as componential analysis. For example, a kinship term might be characterized by gender, age, and degree of relationship. A bird might be characterized by color and size.

There is some debate as to whether such feature analysis is significant to the people using the system or is merely a convenient way to organize information about a classification system.

It is also possible that different features may identified if different informants are asked and that these differences may in fact reflect real disagreements about salient features. Consider the term *brother-in-law* in American culture. Americans generally agree that this is an appropriate kinship category. However, they do not agree on its application in all cases. Consider the following questions that might be posed to an advice column:

ISSN: 2324-9773

Eastman, C. (1994). Anthropological Perspectives on Classification Systems. 5th ASIS SIG/CR Classification Research Workshop, 69-78. doi:10.7152/acro.v5i1.13777

John is married to my sister-in-law? Is he my brother-in-law?
John and Mary are now divorced. Is John's brother David still Mary's brother-in-law?

It is possible to find disagreements among Americans as to the appropriate answers to the questions. These differences appear to derive from different treatments of possible salient features, such as current status of a marriage. Similar disagreements can be found in classification systems in other cultures as well.

STRUCTURES

A semantic field (semantic space) is a group of terms related by meaning which apply within a specific domain. Color terms, biological classifications, and kinship terms are all examples of semantic fields. Semantic fields may also refer to activities or situations:

Baseball: pitcher, base, out, fly, infield
Sewing: needle, thread, patch, darn, scissors
Conference: paper, proceedings, program committee, registration fee

There is generally some organization or structure for the terms in a semantic field. Ethnoscientific studies do not merely identify terms in a semantic field but also address the problem of how they are related to each other. However, the organization as perceived by users of the terms may differ from the organization as determined by an outsider.

Several different organizational structures have been used to describe the structures of semantic fields. These structures are often referred to as folk taxonomies; they may or may not have a hierarchical structure. Possible structures include (but are not limited to) taxonomies, paradigms, and rankings.

A *taxonomy* is a hierarchical structure in which terms are related by inclusion. A biological classification is typically structured as a taxonomy. Systems of color terms can be viewed as shallow taxonomies in which many of the terms are included within others. For example, *maroon* and *scarlet* are both *red*.As discussed earlier, some taxonomies may not be clean and elegant.

A *paradigm* is a classification characterized by two or more dimensions, each of which forms a contrast set of mutually exclusive options. This kind of structure is also referred to as a class product space. Kinship terms are often structured as paradigms, using dimensions such as sex and generation.

An orthogonal space is a paradigm in which the domains are independent, and all possible combinations occur. A nonorthogonal space is one in which not all domains are independent, and some possible combinations do not occur. Both orthogonal nd nonorthogonal spaces have been observed. These dimensions are similar to facets in library classification systems, which may be used independently to classify items.

ISSN: 2324-9773

Eastman, C. (1994). Anthropological Perspectives on Classification Systems. 5th ASIS SIG/CR Classification Research Workshop, 69-78. doi:10.7152/acro.v5i1.13777

A *ranking* is an ordered list of terms based upon perceived value or some other criteria. For example, consider English terms used to describe water temperatures: *icy, cool, lukewarm, hot,* and *scalding*. These terms can be ordered on the basis of temperature.

These simple structures may be combined to form more complex structures. For example, rankings and paradigms may be included within taxonomies.

CATEGORIES

The nature of the categories or classes found in classification systems has been extensively studied across a wide variety of disciplines. Much of this work has been done in psychology. There is still debate over the details of appropriate models.

Categories can be characterized by sets of properties, but there may be no properties which are both necessary and sufficient to characterize a category. Most birds can fly, but some kinds of birds cannot. Adult birds have feathers, but some young ones do not. So neither flight nor feathers is an essential characteristic of birds. Such properties may be treated probabilistically.

Categories also show typicality effects. Some items are regarded as better examples of the category than others. Such central items are sometimes referred to as exemplars or prototypes. For example, robins are regarded by American as more typical birds than chickens are; penguins are even less typical.

Smith and Medin (1981) suggest a category model which contains both probabilistic features and exemplar concepts. Rosch and Mervis (1975) find that the extent to which items are regarded as category members depends on the degree of family resemblance, as indicated by the number of matching features. Lakoff (1984) argues that categories should be viewed as radial structures. Fuzzy models have also been proposed for categories.

Some categories appear to be more fundamental to our classification systems than others. These categories form a cognitively basic level of abstraction and are commonly used in everyday discourse (Rosch, 1978). These basic categories tend to fall at intermediate levels of classification hierarchies and are more clearly distinguished than categories at other levels. *Chair* is a basic category. It occupies an intermediate level in a furniture taxonomy, which contains both more general terms (*furniture*)and more specific terms (*rocker*). The role of basic categories in different kinds of classification systems, including information retrieval thesauri, is discussed in Fernandez and Eastman (1991).

VARIATIONS

Even within the same culture, different individuals who share a classification system in a particular domain may differ in its use. Some of these differences may result from differences in perceived salient features, as discussed in the previous section. Others may result from disagreements as to the appropriate term to use in a particular situation, especially for items not central to a category. For example, two Americans may share a color classification system containing the terms *blue* and *green* but may disagree as to which of those terms applies to a particular shirt. A third source of

ISSN: 2324-9773

Eastman, C. (1994). Anthropological Perspectives on Classification Systems. 5th ASIS SIG/CR Classification Research Workshop, 69-78. doi:10.7152/acro.v5i1.13777

variations is the existence of different versions of a classification system within a culture. The central core of the system may be the same for all users, but there may be variations in the specific names used, in the structure of the classification, or in the number of terms used. For example, a specialist in herbal medicine might use a more elaborate botanical classification than other members of the same culture. In this case, the distinction appears to arise from a difference between experts and lay people.

Boster (1977) discusses the phenomenon of terminological variations and presents data on variations in manioc identification and classification among the Aguaruna, a group of people in northern Peru. (Manioc is a crop used extensively in some areas of the world; it is also the source of tapioca.) Informants were asked to identify manioc varieties in two experimental gardens, one containing 15 common varieties ("easy task") and one containing 61 varieties ("hard task"). The easy task had 53 informants; the hard task had 58 informants. The extent of agreement on terms was determined for each task. The overall agreement (proportion of agreement with the population consensus) ranged from 32% to 78% on the easy task and 16% to 33% on the hard task. Each pair of informants was also compared. Sources of variation included gender, kin relationships, residential groups, and individual expertise. Limited retesting after two months revealed that there was some variation over time even for the same individuals.

There have been a number of other studies addressing terminological variation in a variety of cultures and domains. Berlin (1992) summarizes and discusses a number of studies that involve ethnobiological classifications; both lexical and cognitive variation is observed. These studies are consistent with studies that have been done in the information sciences on variations in choice of index terms, search terms, and subject headings (Saracevic, 1991). Brown (1992) regards terminological variation, at least in the forms of polysemy and synonymy, as a human universal. Such variation is a part of human nature that must be taken into consideration in system design and use. Allen (1991) summarizes some approaches which have developed to accommodate variations in information systems.

METRICS

Various quantitative characteristics of classification systems have been examined. In many cases, the patterns that have been observed are the same as have been observed in bibliometric studies. Some information about sizes and distributions of items within structures is summarized here.

The number and characteristics of generic level classifications in ethnobiological classifications has been extensively examined; these studies are summarized in Berlin (1992). The number of generic level taxa in ethnobotanical classifications of plants ranges from 137 to 956 in the 24 systems compared by Berlin. The mean is significantly lower (197) for nonagricultural groups than it is for agricultural groups (520).The range for 10 ethnozoological classifications compared is 186-606; the mean is 390. Berlin speculates that 500-600 generic taxa represents an upper bound for a manageable system.

The distribution of species within generic taxa has also been examined. The subgeneric taxa can be regarded asa contrast set. Most generic taxa are monotypic; they are not subdivided. A few are

ISSN: 2324-9773

Eastman, C. (1994). Anthropological Perspectives on Classification Systems. 5th ASIS SIG/CR Classification Research Workshop, 69-78. doi:10.7152/acro.v5i1.13777

subdivided.The distribution of species within general in ethnobiological classifications generally shows a Zipfian distribution; this has also been observed in scientific classifications.

Wallace (1962) states that "human folk taxonomies rarely require more than the equivalent of six binary dimensions on any given level of abstraction." Six binary dimensions would result in 64 possible categories. If the dimensions are not binary, one would expect a smaller number. Berlin presents some data on the sizes of 60 subgeneric contrast sets with more than 10 members for 11 ethnobotanical systems. Most are in the range 10-30;only four contain more than 50 members, and all of these are cultivated plants. These numbers are consistent with Wallace's observation.

CONCLUSIONS

Classification and taxonomy are universal approaches to organizing the world. Folk taxonomies exhibit many of the same characteristics and properties as are found in systems constructed for use with information retrieval systems. Classification appears to be a human universal (Brown, 1991).

Most of the systems we have constructed to support retrieval systems are much larger than folk taxonomies; it is to be expected that assistance must be provided to help people cope with systems of this size. This is usually done in current systems.

Variation in term use and classification is to be expected and can not be avoided. Such variation appears to be a human universal. It is clearly not limited to information retrieval applications and is not a consequence of bad or inadequate system design. Systems must therefore be designed to accommodate such variation.

REFERENCES

Bryce L. Allen, "Cognitive research in information science: implications for design," *Annual Review of Information Science and Technology,* Volume 26, Martha E. Williams, editor, published for the American Society for Information Science by Learned Information, Inc., Medford, New Jersey,1991.

Donald E. Brown. *Human Universals*, Temple University Press, Philadelphia, Pennsylvania, 1991.

Brent Berlin, *Ethnobiological Classification: Principles of Categorization of Plants and Animals in Traditional Societies*, Princeton University Press, Princeton, New Jersey, 1992.

Brent Berlin and Paul Kay, *Basic Color Terms: Their Universality and Evolution*, University of California Press, Berkeley, California, 1969.

James Shilts Boster, "'Requiem for the omniscient informant': There's life in the old girl yet," in Dougherty, 1985, pp. 177-197.

Janet W. D. Dougherty, editor, *Directions in Cognitive Anthropology*, University of Illinois Press, Urbana, Illinois, 1985.

Marta J. Fernandez and Caroline M. Eastman, "Basic taxonomic structures and levels of abstraction," in *Advances in Classification Research: Proceedings of the 1st ASIS SIG/CR Classification Research Workshop*, published by Learned Information, Inc. for the American Society for Information Science, pp. 57-68, 1991.

Paul Kay and Chad K. McDaniel, "The linguistic significance of the meanings of basic color terms," *Language*, Vol. 54, pp. 610-646, 1978.

ISSN: 2324-9773

Eastman, C. (1994). Anthropological Perspectives on Classification Systems. 5th ASIS SIG/CR Classification Research Workshop, 69-78. doi:10.7152/acro.v5i1.13777

George Lakoff, *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind,* University of Chicago Press, Chicago, Illinois, 1987.

Eleanor Rosch, "Principles of categorization," in *Cognition and Categorization,* Eleanor Rosch and B. B. Lloyd, editors, Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 27-48, 1978.

Eleanor Rosch and Carolyn B. Mervis, "Family resemblances: studies in the internal structure of categories, *Cognitive Psychology,* Vol. 7, pp. 573-605, 1975.

Tefko Saracevic, "Individual differences in organizing, searching, and retrieving information," *ASIS '91: Proceedings of the 54th ASIS Annual Meeting, October 27-31, 1991*, pp. 82-86.

Ernest L. Schusky, *Manual for Kinship Analysis,* Second Edition, University Press of America, Inc., Lanham, Maryland, 1983.

Edward E. Smith and Douglas L. Medin, *Categories and Concepts,* Harvard University Press, Cambridge, Massachusetts, 1981.

Anthony F. C. Wallace, "Culture and cognition," *Science*, Vol. 135, No. 3501, February 2, 1962, pp.351-357.

ISSN: 2324-9773

Eastman, C. (1994). Anthropological Perspectives on Classification Systems. 5th ASIS SIG/CR Classification Research Workshop, 69-78. doi:10.7152/acro.v5i1.13777

ISSN: 2324-9773