Semantic Relations in Knowledge Organization Systems

Marisa Bräscher

University of Santa Catarina (BR), Information Science Department,
Campus Universitário Reitor João David Ferreira Lima, Trindad, Florianópolis, SC 88040-970,
<marisa.brascher@gmail.com>

Marisa Bräscher is Professor in the Department of Information Science at the Federal University of Santa Catarina, Brazil and works as Colaborator Professor at University of Brasilia, Brazil. Her research interests include indexing , knowledge organization systems and semantic relationships.

Abstract: Semantic relations in knowledge organization systems (KOS) are discussed as well as the need to analyze and systematize the contributions from different areas of knowledge that are devoted to semantic studies in order to collaborate in the definition of a theoretical framework for the study of types of relations included in KOS. Partial results of a survey reveal that, in general, standards and guidelines for developing thesauri are limited to defining and exemplifying types of relationships without guidance concerning the theoretical underpinning of these definitions. The possibilities of a compositional approach to defining the meaning of syntagmatic relations is discussed. Studies on the theoretical foundations that guide the establishment of semantic relations and approaches to be adopted for the preparation of KOS certainly contribute to consolidating a theoretical framework for the area of knowledge organization.

1.0 Introduction

Linguistics, philosophy, psychology, information science and artificial intelligence are some of the areas that deal with different aspects of knowledge representation, which range from studies of human cognition to the representation models artificially constructed for use in specialized systems. The diversity of approaches leads to different concepts about knowledge representation. For the purposes of our work, we consider the representation of knowledge in a limited scope to representations built by man, i.e., models of abstraction of the real world built for a specific purpose. We focus our attention to studies on representations of knowledge that seek to represent the real world through knowledge organization systems (KOS).

These systems awaken particular interest for information science, when they comply with the important role of standardization of terminology adopted for the organization and retrieval of information. The KOS delimits the use of terms and defines concepts and relations in areas of specific knowledge, shared and consensually. The elaboration of KOS involves the analysis of concepts and respective characteristics for the establishment of the position that each concept occupies in a given domain, as well as of its relations with the other concepts that compose a conceptual system.

Dahlberg (1993, 211) defines knowledge organization as "the science that systematically structures and organizes units of knowledge (concepts) according to their inherent knowledge elements (features) and the application of these concepts and classes of concepts organized to objects/subjects." In a more recent work, Dahlberg (2006) distinguishes two applications for the organization of knowledge: the construction of conceptual systems and the correlation or mapping units of this conceptual system with objects of reality. We discuss the semantic relations within the scope of the first application suggested by Dahlberg, considering the importance of relationships in semantic KOS. Therefore, it is necessary to understand the functions and structure of these systems.

176

Knowl. Org. 41(2014)No.2
M. Bräscher. Semantic Relations in Knowledge Organization Systems

2.0 Function and elements of KOS

The KOS are tools used in the process of organizing knowledge. The obtained quality in information retrieval depends heavily on these semantic tools. Zeng (2006) considers the search by subject the most important user task, as evidenced by the resources used by search engines, directories, databases and websites to provide content access. Vickery (1986, 145) notes that the knowledge representation is an ancient research subject in the area of information and depicts the evolution in this area.

Knowledge representation is an issue that concerns documentation since its inception. The problem now is relevant in many other situations beyond documents and indexes. The structure of the records and files of databases, the data structure in computer programs, the syntactic and semantic structure of natural language, knowledge representation in artificial intelligence, models of human memory: in all these fields it is necessary to decide how knowledge might be represented so that these representations can be manipulated. The representation and retrieval of content has been a constant concern in the field of the information science and KOS play an important role in this context. For Hodge (2000), KOS are "mechanisms of organizing information," and, as prefers Vickery (2008), "complementary tools that help the user to find their way into the text." Hodge (2000) even qualifies these systems as "the heart of every library, museum, and archive."

If on one hand there is agreement of the authors on the importance of KOS for the organization and retrieval of information, there does not seem to be much consensus on the scope of the concept of KOS and the types of systems considered. Vickery (2008) indicates indexes and summaries of books as the simplest forms of KOS. However, the author points out that, with time, these systems have become more complex and assumed wider functions, with, as a consequence, the emergence of new denominations, such as taxonomies, classifications, thesauri or ontology. The KOS are schemes that organize, manage and retrieve information, and the basis of modern KOS are terms and indications of relationships between them. We note that the characteristic that Vickery (2008) emphasizes in his proposal is the function of organization and information retrieval that KOS meet at different moments that mark the evolution of information systems. In this respect, we agree with the view of the author and emphasize that information science should occupy mainly of the KOS which fulfill this function.

To arrive at a definition that best represents the interests of information science with regard to the study and development of KOS, we accept that, in principle, these systems are semantic tools, according to the definition proposed by Hjørland (2008): "the KOS are tools that present the organized interpretation of knowledge structures, also called semantic tools. "For the same author, these tools contemplate, essentially, the concepts and their semantic relations, and to inform the meaning of the terms or symbols they employ. We further assume that, for information science, the KOS must fulfill the function of instruments of organization and information retrieval. Given this, we understand that the KOS are representations of knowledge domains that define the meaning of terms in the context of these domains, establish conceptual relations that help position a concept in the conceptual system and are used as instruments of organization and information retrieval.

As Dahlberg (1993, 211) reminds us, the most important aspect of the theoretical basis of the area of knowledge organization is the fact that "any organization of knowledge should be based on units of knowledge—which are nothing more than concepts." On this Hjørland (2007) agrees with Dahlberg and states that "the organization of knowledge is basically organizing concepts." We believe, therefore, that the concepts are essential components of the KOS. To Sowa (1984, 344), the concepts "are inventions of the human mind used to build a model of the world." In the preparation of KOS, the main object of study are the concepts and not the expression used to refer to them (terms, codes, formulas or other symbols that represent them). The Simple Knowledge Organization System (SKOS) calls these expressions of labels or tags (labels), which are also elements of KOS, used to refer to concepts. The SKOS semantic relations assign a crucial role in defining concepts and emphasize that these are not only defined by words of the natural language, but also the links that bind them to other concepts. The semantic relations are thus the third element of the KOS and are associations between concepts. For Green, Bean and Myaeng (2002), concepts are the building blocks of a conceptual framework; the relations between them are the cement that holds them together. Among the three components of the KOS—concepts, terms and semantic relations—we chose the latter as the object of study.

3.0 Semantic relations

Semantic relations are established by analyzing the characteristics or properties of concepts, which allow us to identify differences and similarities that show certain types of relationships. To Khoo and Na (2006), these relationships can be viewed as directional links between concepts and entities that participate in them. Two concepts connected through a relationship can be repre-

Knowl. Org. 41(2014)No.2
M. Bräscher. Semantic Relations in Knowledge Organization Systems

177

sented as a triple [concept] → (relation) → [concept]. A semantic relation contains at least two vacant seats or slots to be filled. A concept that participates in the relationship must have certain semantic features or belong to a certain category of concepts. Most relationships are binary, but there are types of relations that have three or more slots.

The organization of concepts into classes is an arbitrary process, in which we select the relationships to be presented according to the peculiarities of the domain we want to represent. By analyzing a domain, we can group the concepts and relate them in different ways. Sowa (1984, 344) notes that "concepts and perceptions cannot form perfect world models, they are abstractions that select important features for a particular purpose and ignore details or complexities that are important only for certain purposes."

Thus, the same entity can be characterized in different ways, depending on the area in question and the type of knowledge organization system we want to build. The arbitrariness of the selection of groups and relationships cannot be seen, however, as something totally random. It is due to the aspects we want to highlight in a particular knowledge representation. We should pay attention to, above all, as emphasized Hodge (2000), there must be a correspondence between the KOS and the reality they seek to represent.

In a previous study (Bräscher and Carlan 2010), we found that classifications, thesauri, taxonomies and ontologies provide three types of basic semantic relations as defined by ISO 25964-1: hierarchical relation, equivalence relation and associative relation. The equivalence relation is present in all of them through different resources: the remissive *see,* in classifications and taxonomies, relation *use / used* to in thesauri and same as in ontologies. The four types of KOS contemplate the notion of hierarchy among concepts. In thesauri and ontologies, hierarchies follow rigid principles; on the other hand, classifications and taxonomies have no capabilities for accurate differentiation of types of hierarchical relationships. As for this kind of relationship, ISO 25964-1 distinguishes: relations genre/type (generic term/specific term); relations part/whole (generic partitive term/specific partitive term) and instantiation relation (IsA).

In ontologies, the relation genus/species guides the hierarchy class and the principle of inherited characteristics is applied rigorously, so that every member of the subclass inherits the characteristics of the class. In thesaurus, although the rule indicates to the use of distinct symbols for the whole-part relation and to the associative, in certain cases, the part/whole relation and genus-species are treated indistinctively. The part/whole relation as well as other types of relationships between concepts are speci-

fied in ontologies, through the properties defined for classes and subclasses, and the relation of instantiation occurs by the binding of each individual as a member of a particular class or subclass.

The identification of the hierarchical relationship in classifications and taxonomies is made, generally, by the positioning of the terms in the vertical structure, but there is no indication of the specific type of relationship, there is an idea of subordination when categories are expanded, but the relationships within them may be different in nature. When analyzing an example of the Universal Decimal Classification (UDC), there are relationships that can be interpreted by the logic of part-whole, as what occurs between the large area of knowledge expressed by class 6 "Applied Science. Medicine. Technology" and the areas that comprise it (medical sciences, engineering, etc.). In the case of masonry structures (624.012) and concrete structures (624.012.45), the relation is genus/species, since the latter are of the first type. There are still, cases in which the associative relation is presented in the same manner as the hierarchical relation.

Under associative relationships, thesauri consist of all relations other than equivalence or hierarchical. The same does not occur in ontologies, which have the resources to define more precisely the kind of associative relationship: cause/effect, agent/process, and process/tool, among others.

Traditional KOS as classifications and thesauri tend to emphasize paradigmatic relations, but the interest in the representation of syntagmatic relations grows to the extent that today's technology enables the use of these relations in different applications. The distinction between relationships that occur in the paradigmatic and syntagmatic axes was established by structuralist Saussure, and is described in his book published in 1959 (cf. Saussure, 2006). According to this author, the paradigmatic relations occur between units that can occupy the same position in a given sentence. Therefore, these units belong to the same semantic category, such as genus-relations species, whole-part and synonymy. The syntagmatic relations occur due to the linear nature of language and are established among lexical units that precede or succeed each other in the speech.

Green (2008) points out that the set of syntagmatic relations is more comprehensive and its scope is broader than that of the paradigmatic. Perhaps for this reason, the principles governing the syntagmatic relations are more difficult to define. We cannot fail to mention the pioneering studies on semantic relations in the indexing process, as Gardin (1973), and Austin (1974), which resulted, respectively, in systems SYNTOL and PRECIS and a and b (1980a and b), with his method of relational indexing.

4.0 Semantic relations and compositional approach of the significance

KOS provide the structure of relationships that guides the organization and retrieval of information. This structure should therefore reflect particular model of the world and consider the possibilities of semantic relations that effectively assist to represent knowledge. The relations in KOS cannot be too far from the context in which they occur. As affirmed by Khoo and Na (2006), concepts and relationships are defined by language and text, and it is difficult to analyze the meaning of the concepts and relationships without considering language that expresses. In our view, therefore, the linguistic structures found in texts can serve as a source for mapping the syntagmatic relations of a given domain, which may later be represented in KOS.

The KOS generally represent domains of knowledge. The communication in the context of a specific domain, in turn, uses specialized languages. Cabré (1993) reminds us, in the technical and scientific texts, among the components of the sentences, the lexicon is the most important level, especially nominalizations and nominal forms. The specialized languages are characterized also by the strong presence of syntagmatic units of the type name + adjective and name+ preposition + name. The calculation of these compositional units involves analysis of paradigmatic and syntagmatic relations and can contribute to the representation of the meanings in KOS.

We have explored the compositional approach as a possibility to define syntagmatic relations closer to the textual content, which in turn registers knowledge about the real world that we want to represent in KOS. This approach is based on the principle that the meaning of a word can be parsed into smaller units of meaning, called semantic features, sememes or primitives. To Gross (1990) the notion of grammar traditionally rests on the concept of compositionality. This means that the meaning of a sequence is that of the product of its constituent elements. The information processed in the context of semantic grammar refers to the relationship established between units that compose sentences or phrases as compositionality applies to various levels of text segmentation. Thus, a nominal phrase is interpreted according to the meaning of its constituent elements. We are aware that there are expressions which meaning is not that the sum of its constituent parts; spanner wrench is not of English origin, but a kind of tool. In these cases semantic opacity occurs, there is not, therefore, calculation of meaning. For this reason, the opaque expressions should not be factored into KOS, because, otherwise, there will be loss of meaning.

The interpretation of relationships in this type of compositional analysis takes into account the logic of predicates and arguments. From the element that satisfies the predicate function, identifies the number of slots (empty houses) that the unit in the predicate function requires. The valence theory, proposed by Tesnière (1959) can be used to set the number of arguments that a predicate requires. To Tesnière (1959), we cannot express thoughts without connecting concepts; thus, predicates and arguments are necessarily connected by some kind of relationship. In his theory, Tesnière takes the verb as core element, but to Borba (1996), valence has a more generic concept, is the property that has a class of elements being able to connect with other specific classes of elements of which this same property distinguishes this class from others of the same syntagmatic level. The expansion of the valence concept, covering the regime of names and adjectives is also presented by Khoo and Na (2008) by asserting that it is possible that adjectives and names have valence.

In compositional analysis, relations are called themes and are defined by role or function performed by thematic lexical unit in the sentence. The grammar of Fillmore (1968) cases is widely used to define thematic roles. This author defines case as 'a set of universal presumably innate concepts that identify certain types of judgments that human beings are capable of making of the events occurring around him, such as who did it, to whom it happened and what has changed. The first set of cases proposed by Fillmore includes agentive, instrumental, dative, factual locative and objective. The number of cases varies from author to author; Sowa (1984) applies a set of specific cases to support his theory of conceptual graphs. In general, cases are used to indicate the type of relation that occurs between a predicate and its arguments. Borba (1996, 14) thus defines the case:

- syntamatic-semantic relation between a predicate and an argument;
- result of a syntactic relation that has a semantic result;
- underlying category, ie, universal grammar property not to be confused with surface functions like subject, direct object, indirect object such as treaties come in traditional grammar.

For example, we present two cases proposed by Sowa (1984):

- agent - binds an action to an animated, in which the animate represents the actor of the action.
- object - binds an action to an entity that stands the action.

In an analysis of predicate/argument of 'diseases classification,' we have that classification is the core of the predicate, object that has a relation with the entity 'dis-

eases.' In 'notification of justice', the relationship is agent/ process, 'justice' is the notifying. It is necessary to note that the semantic value is not connected to the lexicon itself, but to the relationship established between the constituents of the sentence or phrase. A name has only one agent relation when it relates to another element that has the trait + process that works as the core of the predicate and that, by its valence, requires a certain number and type of arguments.

To establish semantic constraints to the concepts that function as arguments in a structure of predicates and arguments, it is necessary to establish relations in paradigmatic level. Thus, the elements are identified as belonging to certain categories, such as, organization, process, event, animate, inanimate, among others. Certain types of relations only occur between elements of specific categories, the ratio "agent" implies an element of the 'process' class and one of the 'animated' class, for example. The semantic features function as criteria that validate certain relations, so the expressions' the printer ate the paper 'and' I ate a brick, "can only be accepted in a metaphorical sense, because the predicate introduced by the verb eat requires two arguments, which works as an agent must possess the trait + animal and the object must possess the trait + edible.

One of the difficulties in establishing the syntagmatic relations lies in the fact that such relationships are related to the functions performed by their constituent elements. The functions are not intrinsic to the nature of an element: a name can serve as an agent or as an object, perhaps, for this reason the associative relations in the thesaurus allow, in general, for a very broad scope of possibilities, unlike the relations paradigms that are more universal because they are intrinsic or inherent to certain concepts. A dog will always have the characteristic + animated and can only form hierarchy with another genre-kind concept that has this characteristic. But if we consider the possibilities of syntagmatic relations involving the concept of dog, even in contexts defined by specific domains, we shall see they are more difficult to delineate. That certain fragility in the establishment of syntagmatic relations sometimes leads to some inconsistencies in the representation of these relations in KOS. In a thesaurus, for example, synthagmatic relations of any kind are considered associative and the set of associative relations of a given term may vary greatly from one thesaurus to another, even in the same domain.

Perhaps the establishment of syntagmatic relations based on more restricted parameters, supported by linguistic theories already established can help narrow the scope of associative relationships in specific fields. The potential of the valence theory and case grammar should be investigated. As pointed out by Khoo and Na (2008), automatic processing systems of natural language make use of grammar cases to convert the text into a semantic representation. In addition to the applications mentioned in their article, as systems and machine translation systems and question-answer, we investigated whether these semantic representations could support the development of KOS.

5.0 Final thoughts

The issues that are under discussion in the organization of knowledge lead us to believe that thinking about the types of semantic relations and approaches to be adopted for the preparation of KOS certainly contribute to consolidating a theoretical framework for the area of knowledge organization. The advancement of technology opens up new possibilities of exploitation of the different types of relationships we use to express knowledge. The representation of semantic relationships in KOS helps bring closer relations between such models of representation and the reality they wish to express.

In our reflections on semantic relationships in KOS we have realized the complexity of a theoretical approach to the subject, particularly in light of the different contributions found in areas that are devoted to semantic studies. We believe it is necessary to rescue the knowledge gained in decades of research about the indexing languages for the construction of theoretical and methodological framework concerning the KOS, which can use common theoretical and methodological bases, coming from the different areas which work with the representation of knowledge, as in the case of this work, in which we rely on compositional analysis based on valence theory and case grammar. Studies on syntagmatic relationships should be encouraged to contribute to the construction of theoretical reference, helpful to those who are dedicated to the development of knowledge organization systems.

References

Austin, Derek. 1974. The development of PRECIS: a theoretical and technical history. *Journal of documentation* 30: 97-102.

Borba, Francisco S. 1996. *Uma gramática de valências para o português*. São Paulo: Ática.

Bräscher, Marisa and Carlan, Eliana. 2010. Sistemas de organização do conhecimento: antigas e novas linguagens. In Robredo, Jaime and Bräscher, Marisa, eds., *Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento*. Brasília: IBICT, pp. 147-76.

Cabré, María Teresa. 1993. *La terminologia: teoria, metodologia, aplicaciones*. Barcelona: Antartida/Empures.

180

Knowl. Org. 41(2014)No.2
M. Bräscher. Semantic Relations in Knowledge Organization Systems

Dahlberg, Ingetraut. 1993. Knowledge organization: its scope and possibilities. *Knowledge organization* 20: 211-22.

Dahlberg, Ingetraut. 2006. Knowledge organization: a new science? *Knowledge organization* 33: 11-9.

Farradane, Jason. 1980a. Relational indexing: part I. *Journal of information science* 1: 267-76.

Farradane, Jason. 1980b. Relational indexing: part ll. *Journal of information science* 1: 313-24.

Fillmore, Charles J. 1968. The case for case. In Bach, Emmon W. and Harms, Robert T., eds., *Universals in linguistic theory*. New York: Holt, Rinehart, and Winston, pp. 1-88.

Gardin, Jean-Claude. 1973. Document analysis and linguistic theory. *Journal of documentation* 29: 137-68.

Green, Rebecca. 2008. Relationships in knowledge organization. *Knowledge organization* 35: 150-9.

Green, Rebecca, Bean, Carol A. and Myaeng, Sung Hyon. 2002. *The semantic of relationships: an interdisciplinary perspective*. Norwell-USA: Kluwer Academic Publishers.

Gross, Gaston. 1990. Definition des noms composes dans un lexique-grammaire. *Langue Française* 87: 84-90.

Hjørland, Birger. 2007. Semantic and knowledge organization. *Annual Review of Information Science and Technology* 41: 367-405.

Hjørland, Birger. 2008. *Knowledge organization systems (KOS)*. Available http://www.db.dk/bh/lifeboat_ko/CONCEPTS/knowledge_organization_systems.htm

Hodge, Gail. 2000. *Systems of knowledge organization for digital libraries: beyond traditional authorities files*. Washington, DC, the Council on Library and Information Resources. Available http://www.clir.org/pubs/reports/pub91/pub91.pdf

ISO (International Organization for Standardization). 2011. *Information and documentation: thesauri and interoperability with other vocabularies. ISO 25964-1.*

Khoo, Christopher S. G. and Na, Jin-Cheon. 2006. Semantic relations in information science. *Annual Review of Information Science and Technology* 40: 157-228.

Saussure, Ferdinand de. 2006. *Curso de linguística geral*. São Paulo, Cultrix.

Sowa, John F. 1984. *Conceptual structures: information processing in mind and machine*. Massachussetts: Addison-Wesley.

Tesnière, Lucien. 1959. *Eléments de syntaxe structurale*. Paris: C. Klincksieck.

Vickery, Brian C. 1986. Knowledge representation: a brief review. *Journal of documentation* 24: 145-59.

Vickery, Brian C. 2008. *On 'knowledge organisation'*. Available http://www.lucis.me.uk/knowlorg.htm#start

Zeng, Marcia Lei. 2006. Sharing and use of subject authority data. *International cataloguing and bibliographic control* 35 n.3: 52-54.