Hierarchical Relationships "is-a": Distinguishing Belonging, Inclusion and Part/of Relationships.

Christophe Jouis
Université Paris III – Sorbonne Nouvelle, FRANCE
&
LIP6 - Université Pierre et Marie Curie
8, rue du Capitaine Scott, 75015, Paris, FRANCE
cjouis@univ-paris3.fr

Abstract

In thesauri, conceptual structures or semantic networks, relationships are too often vague. For instance, in terminology, the relationships between concepts are often reduced to the distinction established by standard (ISO 704, 1987) and (ISO 1087, 1990) between hierarchical relationships (genus-species relationships and part/whole relationships) and non-hierarchical relationships ("time, space, causal relationships, etc."). The semantics of relationships are vague because the principal users of these relationships are industrial actors (translators of technical handbooks, terminologists, data-processing specialists, etc.). Nevertheless, the consistency of the models built must always be guaranteed... One possible approach to this problem consists in organizing the relationships in a typology based on logical properties. For instance, we typically use only the general relation "Is-a". It is too vague. We assume that general relation "Is-a" is characterized by asymmetry. This asymmetry is specified in: (1) the belonging of one individualizable entity to a distributive class, (2) Inclusion among distributive classes and (3) relation part of (or "composition").

1. Introduction

The semantics of the relationships between concepts (i.e. for each relation, the number and types of its arguments, its algebraic properties, etc.) are often too vague (for example in thesauri, conceptual structure, or semantic networks). One possible approach to this problem consists in organizing the relationships in a typology based on logical properties. For example, (Winton, Chaffin & Herrmann, 1987) or (Pribbenow, 2002) distinguishe various types of part/whole relationships. This typology inspired the treatment of the part/whole relationship in WordNet (Miller, 1990). Recent works applying terminological relationships to information retrieval, in particular to the construction of thesauri, tries to specify the properties of the link between concepts better and to extend non-hierarchical relationships (Molholt, 1996; Green, 1996, 1998, Bean, 1996). Other recent work aims to integrate into the terminological models theories arising from linguistics (semantics, for example) and artificial intelligence, in particular the modeling of knowledge for the design of knowledge-based systems (KBS) and "ontologies", as defined, for example, by (Sowa, 1984, 1996, 2000). In all these disciplines, the need to structure knowledge and then to validate the representations obtained is fundamental. In artificial intelligence, methods for acquiring and modeling knowledge, such as KADSII, as presented in (Wielingua, Schreiber & Breuker, 1992), were developed to assist with the design of KBS. These methods propose modeling a field of expertise in the form of concepts connected by semantic relationships in conceptual object-oriented languages (called "domain level" in KADS). From our standpoint, these languages appear very close to terminological database structures. In terminology, software has been developed to "navigate" networks of concepts structuring micro-domains, for example, the Termisti system (Van Campenhoudt 1994, 1998; Lejeune & Van Campenhoudt, 1998), which considers systems of coherence, the Code system, the Cogniterm project of (Meyer and Mchaffie, 1994), and the Ikarus system of (Meyer and Skuce, 1998), which supports computerized management of terminological knowledge bases.

With the view to better designing the knowledge structures underlying the concepts of a field, and more specifically, the indexing of documents and/or information retrieval, we propose a structured set of relationships, based on a linguistic model, the Applicative and Cognitive Grammar (ACG) of (Descles, 1990). This model was extended to terminology by (Jouis, 1995), and applied by (Mustafa and Jouis, 1996, 1997) and by (Jouis 1998, 2004).

2. Proposed semantic and logical system

Relations are a part of a specification network based on a general terminological scheme (i.e., a coherent system of meaning of relations). In such a system, a specific relation may be characterized as to its: (1) functional type (the semantic type of argument of the relation); (2) Algebraic properties (reflexivity, symmetry, transitivity, etc.); and (3) combinatorial relations with others entities in the same context (for instance, the part of the text where a concept is defined). We distinguish four categories of primitive:

- **Elementary semantic types of entities.** We distinguish a certain number of elementary types of entities. For instance: Boolean entities (noted H) are objects, whose value is either true or false. Individualizable entities are entities that can be designated and shown by pointing. They may be counted individually or regrouped by distributive classes. Entities such as *John, table, chair, man, child* are distinctive. Individualizable entities are noted J: [J: table]. Distributive classes regroup individual entities with one identical property. They are noted D. For instance: [D: to-be-a-square]. Collective classes are distinguished from Distributive classes in that they represent objects that form a "whole" from more elementary objects. They are noted C. Thus, [C: geographical entities], [C: molecule] represent collective classes. The "whole" is seen as the "accumulation" of elements that constitute it, disjoint or not. Lesniewsky (1886-1939) proposed a general theory of wholes and parts (mereology), in response to the problem of set theory (Cantor, 1932, 1962). A detailed analysis of mereology was carried out by Miéville (1984). Lesniewsky arrives at the conclusion that the notion of class contains two features: the distributive one and the collective one. The following example, borrowed from (Grize, 1973) give an idea of the difference: "A distributive class is, to be strictly correct, the extension of a concept. If p is the concept planet, the statement that *Jupiter is a planet* is either to pose p(Jupiter) or Jupiter $\in \{x \mid p(x)\}$, and the transmitted information is the same one in the two writing. Thus p = {Mercury, Venus, Earth, Mars, Jupiter, Saturn, Neptune, Pluto} is a distributive class. It contains nine elements and nothing else. The polar caps of Mars, the red Jupiter spot, the rings of Saturn do not belong to p. Yet, all that and a thousand other things deal with the concept planet. The notion of collective class must mitigate this gap" (see figure 1).

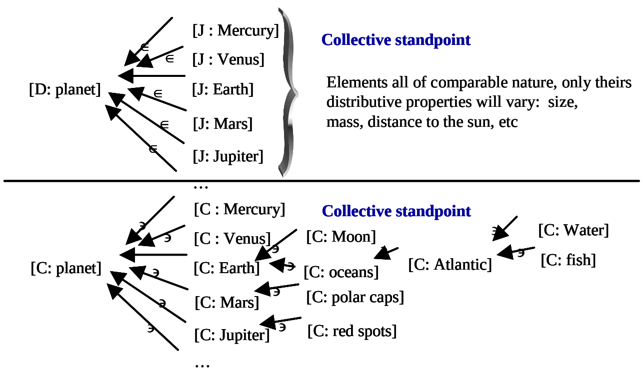

Figure 1 : Distributive vs Collective (part-of) classes: different but logical and coherent standpoint

- **Formation operators**, which create more complex types from elementary types (lists, arrays, functional types[1], etc.). For instance, it is possible to define functional types. From the set of elementary types S = {H, J, D, C, …} we define a system of more complicated types in a recursive way starting from the following rules :

(1) The elements of S are elementary types.
(2) If x and y are types, then Fxy is a functional type.

Then, an entity E of type Fxy (noted [Fxy:E]) is a unary operator which takes for its argument an object of the type x to provide a result of the type y. If we consider an entity A of type x, the application of E to A will build a certain entity B of the type y: ([Fxy: E] [x: A]) > [y: B]. For example, type FJH is that of an operator which, when applied to an individualizable entity (J) returns a value of truth H (set of individuals, or "concept" such as [FJH: "to-be-a-square"]). A relation between an individual entity and a distributive class will have type FJFDH. Because this relation is a binary operator, the application is done in two steps. For example, we have the following types: [J: Jean], [D: Human] and [FJFDH: inclusion]. The inclusion[2] applies initially to Jean to return an operator of the type FDH:

([FJFDH: inclusion] [J: Jean]) >
[FDH: inclusion_Jean]. The result is an operator of type FDH that applies to the distributive class Human to return a value of truth of type H:

([FDH: inclusion_Jean] [D : Human] >
[H : True].

All representations of the cognitive level are typified in this manner.

- **Fundamental static relations between entities**, where static relationships enable the description of some states related to an area of knowledge. We have identified twenty static relationships. They are structured and independent from a particular domain. Static relationships are binary relations. In this paper, we will describe more specifically belonging, inclusion and part/whole relationships, because they are related to the general "Is-a" relationship.

- **Fundamental dynamic relations** between terminological units, where dynamic relationships enable the description of processes or events related to an area of knowledge: movements, changes of state, conservation of a movement, iterations, intensity, variation, constraints, causes, etc.

Relationships are therefore classified in two main, disjoint categories: static relations and dynamic relations. In this paper, we will describe more specifically the static relations, and in particular those related to the general "Is-a" relationship.

[1] In the meaning of the Church typed "Lambda-calculus" theory or typed logic theory of Curry (Curry & Feys, 1958).

[2] The relation of inclusion is defined more formally in part 3.

3. Specification of "Is-a" relationships

In our system, a relation may be specified in more precise relations in terms of its properties. We assume that general relation "Is-a" is characterized by asymmetry. This asymmetry is specified in:

- The belonging of one individualizable entity to a distributive class (noted $\in$). Of type FJFDH, this relation is NEVER-reflexive, asymmetric and NEVER transitive[3]. It is expressed in statement such as: *π is a real*.
- Inclusion among distributive classes, noted $\subset$, (e.g. *Bacteria are microorganisms*), which is of type FDFDH, is NEVER-reflexive, asymmetric and transitive. It should be noted that, in many thesaurus or semantic network models, we typically use only the general relation "Is-a" without distinguishing belonging from inclusion. However, there is a fundamental difference, since the first is NEVER-transitive while the second is transitive and allows inheritance of properties.
- The relation part of (or "composition"), noted $\ni$, is reflexive but (generally) non transitive. It is expressed in statement like *The hand forms part of the arm*. Its type is FCFxH, where x is of type J or of type C. Part of is specified in several relations. Indeed, there are a great number of properties describing the relationships between the composing object and the total object (collective class), for example:
    - Atomic composition versus non-atomic composition (*The smallest component of a program is the bit* versus *A book breaks up into chapters, which themselves break up into paragraphs*). Atomic composition does not admit transitivity, but atomic composition authorizes it.
    - Direct composition versus non direct composition (*Opium appears among the primary component of Lamaline* versus *A molecule consist of neutrons, protons and electrons, which are part of atoms*). An Object-Part OP is a direct component of the Object-Whole OW, if there is no object OP1 (different from OP) such that object OP is a component of object OP1 and object OP1 is a component of object OW Otherwise, OP is a non-direct component. (see figure 2). Non direct composition is transitive, while direct composition is non transitive.

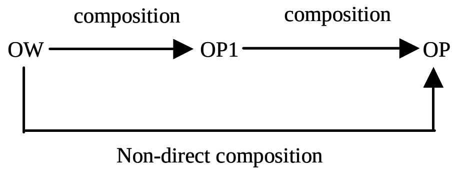

Figure 2. Direct vs. non-direct composition

    - Necessary composition versus non-necessary composition (*The processor is one of the essential components of a computer* versus *A CD-ROM drive is an accessory component of a computer*). The characteristics necessary and non necessary are transitive within the relation of composition.
    - Single composition versus non single composition (*A young star is made up exclusively of atoms of hydrogen* versus *The atmosphere is a mixture of several gases, whose principal one are oxygen and nitrogen*).
    - Quantifiable composition versus non-quantifiable composition (*The hand is made up of five fingers; Each human cell contains 46 chromosomes* versus *Water consist of atoms of oxygen and atoms of hydrogen*).

Acknowledgements

This research is sponsored by the LIP6 (Laboratoire d'Informatique de PARIS 6 – Université Pierre et Marie Curie, France). Particularly, the author thank Mr. Pr. Jean-Gabriel Ganascia, in charge of the team ACASA ("Acquisition de Connaissances et Apprentissage Symbolique Automatique »). He helped the author to better formalize the model presented in this paper.

References

Bean, C. (1996). Analysis of non-hierarchical associative relationships among Medical Subject Headings (MeSH), Anatomical Terminology, Knowledge Organization and Change: Proceedings of the Fourth International ISKO Conference, 80-86.

Cantor, G. (1962) Gesammelte Abhandlungen (E. Zermelo, Ed.) Hildesheim. Omls (Original work published 1932).

Curry, H. & Feys, R. (1958). Combinatory Logic (Volume I). North-Holland.

Descles, J.-P. (1990). *Langages Applicatifs, Langues Naturelles et Cognition*. Paris: Hermès.

Green, R. (1998). Attribution and relationality. Structures and Relations in Knowledge Organization: Proceedings of the Fifth International ISKO Conference, 328-315

Green, R. (1996). Development of a relational thesaurus. Proceedings of the Fourth International ISKO Conference, 72-79.

Grize, J.-B. (1973). Logique moderne (Fascicule II). Paris: Mouton/Gauthier-Villars.

[3] We point out in particular the following properties of transitivity for binary relation:
Given three entities X, Y and Z, and a relation R:
R transitive $=_{def}$ $\forall$X, Y, Z R(X, Y) and R(Y,Z) $\Rightarrow$ R(X, Z).
R non transitive $=_{def}$ $\exists$X, Y, Z R(X, Y) and R(Y, Z) and NOT(R(X, Z)).
R NEVER transitive $=_{def}$ $\forall$X, Y, Z (R(X, Y) and R(Y, Z)) $\Rightarrow$ NOT R(X, Z).

International Organization for Standardization (ISO). (1990). *Terminology – Vocabulary = Terminologie – Vocabulaire*. Genève: Organisation internationale de normalisation (ISO 1087 – 1990.).

International Organization for Standardization (ISO). (1987). *Principes et Méthodes de la Terminologie*. Genève: Organisation internationale de normalisation (ISO 704 – 1987.).

Jouis, C. & Ferru, J.-M.(2004). « Intranet Try To Find Project (ITTF): An approach for the searching of relevant information inside an organization », *LREC 2004: Language Resources and Technology Evaluation within Human Language Technologies*, pp. 1325-1329, ELRA – European Language Resources Association, Lisbon, may 2004.

Jouis, C, (1995). « SEEK, un logiciel d'acquisition des connaissances u*tilisant un savoir linguistique sans employer de connaissances sur le monde externe ». In Actes des 6ème Journées Acquisition, Validation, (JAVA 95)*, INRIA, pp. 159--172, Grenoble, Avril 95.

Jouis, C. (1999) "System of Types + Inter-Concept Relations Properties: Towards Validation of Constructed Terminologies", *In Proceedings of the Fifth International ISKO Conference on "Structures and Relations in Knowledge Organisation"*, Lille (France) August 25-29 1998, organized by the International Society for Knowledge Organization & the University Charles de Gaulle Lille 3, Advances in Knowledge Organization, Vol. 6, Editions ERGON Verlag, Frankfurt/Main, pp. 39-47.

Lejeune, N. & Van Campenhoudt, M. (1998). Modèle de données et validité structurelle de fiches terminologiques; L'expérience des microglossaires de TERMISTI. In Blanchet (Ed.) La banque des Mots: Terminologie et Qualité, numéro spécial 8, 97-111.

Miéville, D. (1984). Un développement des Systèmes Logiques de Stanislaw Lesniewsky : Prothétique, Ontologie, Méréologie. Bern. Frankfurt am Main, and New York P. Lang.

Miller, G.A. (1990). Nouns in WordNet: A lexical inheritance system. International Journal of Lexicography, 3, 245-264.

Molholt,P. (1996) Standardization of interconcept links and their usage, Knowledge Organization and Change: Proceedings of the Fourth International ISKO Conference, 65-71

Mustafa, W. & Jouis, C. (1998) "Terminology Extraction and acquisition from textual data: criteria for evaluating tools and methods " In *Proceedings of the First International Conference On Language Resources and Evaluation, Granada (Spain)* : 28-30 May 1998, organized by ELRA (European Language Resources Association). Granada: ELRA, Vol. 2, pp. 1175-1180.

Mustafa, W., Jouis C. (1997), "Natural Language Processing-based Techniques and their Use in Data Modelling and Information Retrieval", In *Proceedings of 6th International Study Conference on Classification Research, Knowledge Organization for Information Retrieval*, 16-19 June 1997, University College of London, London, FID/CR, & ISKO. The Hague: FID, pp. 157-161.

Mustafa, W. & Jouis, C. (1996), "Natural Language Processing-based Systems for Terminological Construction and their Contribution to Information Retrieval", in *Proceedings of the Fourth International Congress on Terminology and Knowledge Engineering (TKE'96)*, Vienna, Austria, INDEX Verlag, Frankfurt/Main. 118-130.

Pribbenow, S. (2002), "Meronymic Relationships: From Classical Mereology to Complex Part-Whole Relations", in The Semantics of Relationships, An Interdisciplinary Perspective, Chapter 3, pp. 35-50, R. Green, C. A. Bean, Sung Hyon Myaeng (Eds), Kluwer Academic Publishers: Dordrecht/Boston/London.

Sowa, J. F. (1984) Conceptual Structures: Information Processing In Mind And Machine, Reading, Mass.: Addison-Wesley.

Sowa, J. F. (1996, August). Ontologies for knowledge sharing. Paper presented at Terminology and Knowledge Engneering'96, Vienna, Austria.

Sowa, J. F. (2000). Knowledge Representation: Logical, Philosophical, and Computational Foundation. Pacific Grove, CA: Brooks/Cole.

Van Campenhoudt, M. (1998). Abrégé de Terminologie Multilingue [On-line]. Available : http://www.termisti.refer.org/marcweb.htm> [2006, junuary 4]

Van Campenhoudt, M. (1994). Les relation notionnelles expérimentées dans les microglossaires de TERMISTI : Du foisonnement à la régularité. In A. Cla & P. Bouillon (Eds). TA-TAO. : Recherches de Pointe et Applications Immédiates. Troisièmes Journées Scientifiques du Réseau Thématique de Recherche « Lexicologie, Terminologie et Traduction », 409-423 Montréal and Beyrouth : AUPELF-UREF & F.M.A.

Wielingua, B., Schreiber, A., Breuker, J., (1992). KADS: A modeling approach to knowledge engineering? Knowledge Acquisition, 4 (1), 5-53.

Winston, M.E., Chaffin, R., Herrmann, D (1987). A taxonomy of part-whole relation. Cognitive Science, 11, 417-444.