

# Research Trajectories in Knowledge Organization

Series Editor: María-José López-Huertas

## Indexing Languages and Subject Representation: A Fascinating and Inexhaustible Field of Research

Michèle Hudon

Université de Montréal, École de bibliothéconomie et des sciences de l'information,  
C.P. 6128, succursale Centre-ville, Montréal, QC H3C 3J7, Canada,  
<Michele.hudon@umontreal.ca>



Michèle Hudon was Professor at École de bibliothéconomie et des sciences de l'information (EBSI), Université de Montréal (UdeM), from 1994 to 2016. She holds a master's degree in library science from Université de Montréal, and a PhD in information studies from the University of Toronto. Prior to joining the faculty at UdeM, she worked as bilingual indexing and thesaurus specialist for the Canadian and Québec governments, as well as with several non-profit organizations. She is the author of *Guide pratique pour l'élaboration d'un thésaurus documentaire* (ASTED 2009), of *Analyse et représentation documentaires : introduction à l'indexation, à la classification et à la condensation des documents* (PUQ 2013), and of several papers on indexing languages (development and application), on classification and the virtual library, and on teaching classification.

Hudon, Michèle. 2016. "Indexing Languages and Subject Representation: A Fascinating and Inexhaustible Field of Research." *Knowledge Organization* 43: 550-562. 42 references.

**Abstract:** The author looks back on three and a half decades of professional, teaching, and research work in the field of knowledge organization. As a researcher, she gave herself three objectives: 1) to make available to professionals and students, in the form of analyses and reviews, the results of research in the fields of subject analysis and indexing languages development and application; 2) to generate observations, questions, and results of immediate applicability; and, 3) to contribute to the discussion of important issues in knowledge organization and in library and information science more generally. This article is structured around seven themes, each corresponding to a project conducted as independent researcher or as part of a team.

Received: 24 August 2016; Accepted: 1 September 2016

Keywords: classification, structure, terms, thesaurus, thesauri, conceptual level, facets

### 1.0 Introduction

The invitation to submit a paper that would summarize my career as a researcher in the field of knowledge organization (KO) came in the fall of 2015, just as I was contemplating retirement. The proposal was challenging; it meant looking back at three and a half decades of professional and research work, and reading again, from a perspective altered by time, a selection of papers published over the years to present the results of this work.

My interests as library and information science (LIS) master's and doctoral student, as professional librarian, as

LIS independent consultant, and as LIS professor have always revolved around indexing languages (ILs) and subject representation. As the following pages will show, I was mostly interested in thesauri and classification schemes, in their essential characteristics, in their design, and in the role they play in preserving the quality and interoperability of information systems.

My research cannot be classified as theoretical. I joined the ranks of teachers and researchers but always kept close to LIS professionals, regularly acting as consultant, trainer, developer and evaluator in various subject access projects. As a researcher, I gave myself three ob-

jectives: 1) to make available to LIS professionals and students, in the form of analyses and reviews, the results of academic research in the fields of subject analysis and IL development, maintenance and application; 2) to generate observations, questions, and methodologically valid results of immediate applicability; and, 3) to contribute to the discussion of problematic issues in KO and in LIS more generally.

This article is structured around seven themes, each corresponding to a project conducted as independent researcher or as part of a team. Each project generated several publications in journals, professional and academic, and in conference proceedings; references to the most complete papers, where a full description of methodology and detailed results will be found, are provided in each section.

## 2.0 Background

My interest in the process of subject representation and in LIS dates back to my years as an undergraduate student in literature and linguistics, when I was offered a part-time job as indexer for *RADAR* (*Répertoire analytique d'articles de revues*), then the primary index providing access to professional and academic journals published in Québec. Over three summers, I also had the opportunity to work at the Assemblée nationale du Québec's library, then in the process of completing the colossal task of re-constituting nineteenth century parliamentary debates; I was, of course, tasked with the production of detailed name and subject indexes. These "student" jobs, for which I developed an immediate passion, would lead to career paths I could not have envisioned till then. As a LIS master's student, I dutifully selected all classification and indexing courses that were offered, and decided that my future was in technical services if I were to work in a library. This I did, in a small university, but for a few years only.

I don't remember how and why I was offered a position as lecturer, and then as regular full-time faculty, in the Library Techniques program of the local Cégep<sup>1</sup>. Although I had not considered teaching as a career, I was no doubt attracted to the fact that I would be responsible for all courses in the areas of cataloging, classification, and indexing. Teaching represented the best way to increase my knowledge of these fields, and it is with little apprehension that I started on this new career course.

Although I continued to teach in several library techniques programs over the following 15 years, the first teaching assignment led quickly to a desire to know and to do more, to explore the LIS literature, to dig into theory, to test hypotheses and to write about experimental results. Doctoral studies would allow me to do this, and would eventually lead to another career switch, this time

to a position as university professor and researcher at Université de Montréal.

### 3.0 Thesauri and definition

Having maintained over the years a strong interest for thesauri, discovered while I was in library school, my doctoral research focused quite naturally on this type of IL. Since I was also curious about terminology and its methods, considering the thesaurus to be at the junction of information science (IS) and terminology (Hudon 2006), I attempted to link thesauri and terminology through the concept of terminological or standardized definition (Hudon 1996, 1998a). I believed that such definition would add useful semantic information to that already available to the indexer at the time of descriptor selection.

A main function of the thesaurus is that of clarifying meaning. When the meaning of thesaurus descriptors is unambiguous, subject analysts are expected to be more consistent in their selection of index terms. In the thesaurus, semantic relations and scope notes have traditionally been the main sources of definition, but the low levels of terminological consistency achieved by indexers suggest that these means of providing semantic information are not always efficient. Inconsistency in descriptor assignment affects negatively representational predictability, indexing correctness, and overall indexing quality.

Analytical definitions, which delineate the meaning of a concept by identification of its *genus proximus* and *differentia specifica*, are the foundation of terminological systems. Definitions by synthesis, i.e. through relations, were not viewed as a strong asset of thesauri by Sager (1990), who considered it simplistic to assume that concepts could be adequately defined by three types of relations only (generic, partitive, other). He thus proposed (Sager 1982, 1990) to develop terminological thesauri in which each descriptor record would include a domain dependent standardized definition describing the intension of the concept and determining its position in the appropriate concept system.

Inspired by Sager's proposal, my doctoral research, supervised by Professor Nancy J. Williamson, was conducted in two phases. Phase 1 consisted in the creation of a prototype thesaurus including standardized definitions, the utility of which would be assessed during Phase 2.

The creation of the prototype thesaurus involved the modification of a simple defining model proposed by Sager and L'Homme (1994). Whereby the researchers included as their sixth defining element (out of 7) the essential distinguishing characteristics of the *definiendum*, my own model split element 6 into seven distinct types of characteristics: eCI: [nature: being ...], eC2: [purpose/function], eC3: [means/instrumentation], eC4: [origin], eC5: [destina-

tion], eC6: [place], eC7: [time]; a few years later, I would likely have called these facets!

Three hundred sixty-seven descriptors, along with 243 synonyms, were extracted from the existing *Canadian Literacy Thesaurus / Thésaurus canadien d'alphabétisation* (<http://www.thesaurusalpha.org/>). The conceptual structure and terminology of the field of adult literacy provided several examples of the semantic ambiguity characteristic of the social sciences. A standardized definition was written for each descriptor, with the help of the pre-established defining template.

Phase 2 of the research consisted in a controlled experiment designed to study the effect on inter-indexer terminological consistency of modifying the type and volume of semantic information provided with descriptors in a thesaurus. Three versions of the prototype thesaurus were used: a control version without definitions, an augmented version with definitions and full relational network, and a stripped version with definitions but without associative relations. Novice indexers, randomly assigned to one of three test groups, indexed the same collection of informative abstracts using one version of the prototype thesaurus. Hooper's indexer-pair formula (Hooper 1966) was used to calculate consistency; group consistency figures were obtained by averaging indexer-pair results for each document indexed by all participants in the study, and *t* tests were run for statistical significance.

The reasoning behind the research question and hypotheses was that, when offered the type of prescriptive meaning information that a terminological definition can provide, indexers would not need suggestive associative relations to help them decide whether a descriptor should be assigned or not. A reduction in the number of descriptors used, with evidence of comparable consistency levels in a group of indexers having no access to associative relations and in a group of indexers having access to them, would be interpreted as positive effects of the availability of standardized definitions at the time of descriptor selection.

The study also provided insights into the respective usefulness of standardized definitions and of traditional networks of hierarchical and associative relations as means of specifying meaning in the thesaurus used as indexing aid. The most significant finding was that indexers who did not have access to semantic relations were significantly less consistent in their use of descriptors. The assignment of fewer descriptors by indexers using the stripped thesaurus, rather than being beneficial to consistency, appears to have had the opposite effect; indexers used fewer terms, but they used different terms to represent subject content. The answer to the original research question was that the availability of standardized definitions in a thesaurus was not enough, in the absence of a display of semantic relations, to lead to the same levels of consistency as that achieved

by indexers working with a standard thesaurus, at least where indexers were novices and non-specialists.

At the time of the research, many organizations were making documents widely available on the internet, and building complex intranets for their internal operations, exacerbating an access problem which no doubt existed before without being critical. These organizations needed a controlled list of terms and/or categories to structure this mass of information, but were reluctant to get into IL construction. A solution could have been a new type of tool, such as the experimental stripped thesaurus, which offered basic control of synonyms and meanings, without involving huge costs in development and maintenance. By the time it became possible for me to pursue this line of research, however, the technology had evolved, Google was working its magic, and taxonomies and folksonomies were coming to mind much faster than thesauri when subject access was considered.

#### 4.0 Thesauri and cultural issues

Before and during my doctoral studies, I worked as chief editor of several thesauri, among which the *Canadian Education Thesaurus/Thésaurus canadien de l'éducation*, the aforementioned *Canadian Literacy Thesaurus/Thésaurus canadien d'alphabétisation*, and the *Government of Canada Core Subject Thesaurus/Thésaurus des sujets de base du gouvernement du Canada* ([www.thesaurus.gc.ca/](http://www.thesaurus.gc.ca/)). As made clear by their titles, these ILs and most others I helped develop were bilingual and bicultural. It always appeared vital to me that both Canadian cultures needed to be given equal status in these thesauri, and that French and English should alternate as source and target languages. This position was never directly linked to a specific research project, but it did lead to years of reflecting on the subject, to several contributions to the literature (Hudon 1997, 1998b, 1999, and 2001a), and to active participation in the preparation of ISO25964 *Thesaurus and Interoperability with Other Vocabularies*, published in 2011 and 2013 (Hudon 2012).

With the expansion of the web, the language barrier became a more critical issue than it had ever been. Multilingual thesauri, offering full conceptual and terminological inventories as well as a fully developed relational structure for each natural language represented, were called upon to facilitate networked information communication and retrieval.

Multilingualism in information systems had become a preoccupation in Europe in the 1960s. Multilingual thesauri were quickly developed in various fields (labour, education, health sciences, etc.) and software was designed to facilitate the task. This software provided for the creation of a monolingual structure, usually in English, and then generated automatically other language versions using

a simple file of equivalents. In such thesauri, strong compatibility resulted from full correspondence of concepts, terms and relations, while weak compatibility resulted from correspondence between concepts but not necessarily between terms and relations. As a main concern, practical applicability was always preferred to conceptual correctness.

Over the years, three significant problems came to be associated with multilingual thesauri: 1) that of stretching a language to make it fit a foreign conceptual structure; 2) that of transferring a whole conceptual structure from one culture to another; and, 3) that of translating literally source terms into meaningless expressions in the target language. Such problems stemmed from the dismissal of the fact that there was more to multilingual thesaurus development than finding term equivalents and that there were definite cultural and political dimensions to the process. Languages are much more than lists of words and sets of rules to combine them; they are organized conceptual and lexical structures reflecting the way their speakers see and interact with the real world. A word or term covers a certain area in a conceptual space, and this area can and will vary in different languages. This obviously makes it difficult to translate one natural language into another. In the process of constructing the multilingual thesaurus, this leads to well documented difficulties in determining inter-language equivalence, an operation at best delicate and often controversial; in the multilingual thesaurus, equivalent descriptors must have equivalent connotations.

In multilingual thesauri, if descriptors set as equivalents do not necessarily refer to the exact same concept or cover the exact same area in the conceptual space, it seems obvious that relations between concepts within languages will also vary. A decision has to be made regarding identity and symmetry of semantic structures in the different language versions. There are two views on the matter. The most common view is that all language versions must be identical and symmetrical; each descriptor must have one and only one equivalent in a target language (no single-to-multiple equivalence is allowed), and be related to the same terms and only to these terms. This artificializes all the languages involved, by forcing equivalence where it does not exist (when one source concept/term = no target concept/term), by ignoring true equivalence where it does exist (when one source concept/term = two or more target concepts/terms), by generating incorrect or illogical hierarchies (when a concept/term belongs to a hierarchy in the source language, but to a different hierarchy in the target language), etc. Another view, prevalent in minority cultures, advocates nonidentical and nonsymmetrical structures. The number of descriptors in different language versions must be allowed to vary; concepts which exist in a culture are represented in its language, but if those same concepts do not exist in another

culture, it is unlikely that equivalent verbal representations will be available. Paradigmatic links, hierarchical relations for example, are not necessarily valid in all languages. When two top terms or broad terms are inexact or partial equivalents, they may have a slightly different extension, and consequently different subordinate terms. A multilingual thesaurus in which the relational structures are allowed to differ from one language version to another is more likely to faithfully represent different universe of concepts.

Of the three standard approaches to developing multilingual thesauri (translation of a monolingual thesaurus, merging of several monolingual thesauri, simultaneous development of distinct language versions), the third one offers the stronger guarantees for equal treatment of each natural language. Building multilingual thesauri from the ground up, in complete respect of all languages and cultures involved, will ensure the production of an IL that better reflect the conceptual and terminological structures with which potential users are most familiar. Software must allow for the rotation of source and target languages and for the creation of distinct records for descriptors in every language used in the thesaurus, tolerating conceptual and relational differences within a compatible structure.

I proposed that true equality for distinct languages and cultures represented in a multilingual thesaurus had a better chance of being achieved if the following requirements were met:

- the thesaurus was built within a semi-centralized administrative structure, with representatives of each language/culture on the decision-making team;
- all language versions of the thesaurus were developed simultaneously from the ground up;
- the thesaurus designers were native speakers of the language in which they worked, with a good knowledge of several other languages;
- distinct termbanks were built independently for each source language with terms found in source language documents;
- identity and symmetry of structures were not required across the distinct language versions of the thesaurus, and single-to-multiple equivalence, “orphans,” and variations in hierarchies, etc. were allowed;
- software which allows for non-identity of descriptor records and for rotation of source and target languages was used;
- distinct complete displays for each natural language represented in the thesaurus could be generated.

## 5.0 Indexing languages and moving images

I am definitely a “text” person. Text documents, in print or digital format, have consistently been at the core of my

work. As a member of a small team led by EBSI's Professor James M. Turner, however, I briefly explored the world of non-art moving images, of which my colleague was an expert. We completed three distinct projects, which strengthened my interest in matters of compatibility and interoperability.

The first and main project was designed to study controlled vocabularies used to index non-art moving images<sup>2</sup>; its framework was the high-level stream of access to images, characterized by the use of text to generate information useful for retrieval but not available on the image itself. The world of moving image collection organization was one of locally established techniques, with little or no standardization and little concern for interoperability; this was becoming a problem in a networked environment within which isolation and self-sufficiency were not anymore viable options. Our data (Hudon, Turner and Devin, 2000, 2001) confirmed that the organization and exploitation of moving image collections remained heavily dependent on *ad hoc* information systems structured around locally designed methods and tools. Our assumption that the number of terms needed to describe the contents of everyday film and video materials would level off at a certain point was supported by film and television librarians who were managing their own ILs. Our research questions were:

- How many terms must an IL contain to describe adequately a general collection of everyday images?;
- Are these terms the same from one collection to the next, or are collections so particular that locally maintained tools remain essential?; and,
- Is it reasonable to think that a general thesaurus naming categories of persons, everyday objects and events could be shared among image collection managers?

Such a thesaurus would likely improve the quality of indexing and retrieval within an organization, while considerably increasing the compatibility and interoperability of moving image retrieval systems.

Eleven organizations, managing among themselves a total of 14 distinct collections, provided quantitative and qualitative data for analysis, through a questionnaire and structured interviews. They were television networks and movie production studios on both sides of the Canadian-American border. To provide subject access to their collections, they used keywords, classification schemes, commercial or in-house thesauri, and subject headings. Seven ILs were made available to us, in full or partially, for conceptual and lexical analysis.

An important proportion of terms (almost a third in each IL) were names (of individuals, of institutions, of places). We noted the small number of lead-in terms (non-descriptors), and deduced that the control of synonyms

and identification of conceptual equivalence had not been effected, thus weakening the utility of the tool for indexing and retrieval. We observed the very broad range of domains represented, an obvious necessity to respond to the needs of the varied clienteles of television networks and production studios.

To characterize the ILs, a sample of 2292 distinct terms was generated; it listed all terms starting with the letters F, I, and R in the seven thesauri, minus terms that were proper names. 1838 terms, or 81% of the sample, represented concrete entities, e.g. a cat, a chair, a flower; they would be used to index images at Panofsky's pre-iconographical level and at Shatford's offness level. The most interesting finding was that there was little overlap in the seven vocabularies; seven terms only were present in all ILs, while 1630 terms, or 73% of the sample, showed up in one IL only. Given the presumed similarity of contents represented in the collections managed by participating organizations, these results were surprising. Furthermore, they did little to support our hypothesis that the number of terms needed to index general collections of non-art moving images might be limited, and that these terms would already be present in existing ILs. But the finding needed to be interpreted in light of the fact that more than likely, if we had been in a position to make a full comparison of concepts and terms used in all ILs, we would have found a large number of synonyms among the descriptors with a frequency of 1. It was suggested that a more complete analysis, which would look at concepts and not just at lexical forms, was needed before we could conclude that these ILs were incompatible.

In a follow-up project (Hudon 2004), I pursued this task with the objective of estimating levels of conceptual redundancy in these ILs; common concepts could be used as a basis in the development of a general controlled vocabulary usable for representing categories of persons, objects and events depicted in general collections of non-art moving images. A secondary objective was to test the efficiency of a simple methodology for estimating conceptual compatibility in controlled ILs more generally.

Compatibility of ILs is measured at one or more of four levels. Lexical compatibility exists between terms, e.g. farm houses and farmhouses; conceptual compatibility goes beyond terms to uncover similarities and differences in concepts represented, e.g. farm workers and farm labourers; structural compatibility is to be found in the network of relations between concepts; subject compatibility refers to the possibility for two or more ILs to represent the same subject, whether by means of a single descriptor or by a combination of terms (this latter type defined by Riesthuis 1996). This project focused on lexical and conceptual compatibility, while indirectly looking also at subject compatibility.

The term-to-term comparison method was used to identify types and levels of conceptual equivalence among the five thesauri to which we had been given full access. Six degrees of equivalence were defined: exact lexical equivalence, exact conceptual equivalence, exact equivalence through combinations, partial equivalence through hierarchy (narrow to broad or broad to narrow), partial equivalence through association, and non-equivalence. As expected, conceptual compatibility, i.e. the total of exact conceptual equivalence, exact equivalence through combinations, and all cases of partial equivalence, was significantly higher than exact lexical equivalence. Partial equivalence through hierarchy (narrow to broad), e.g. Thes1 ragweed, Thes4 weeds, was the most frequent case of conceptual equivalence; this suggested that the most significant difference between ILs was a difference in specificity rather than a difference in coverage. Overall, and when taking all types of equivalence into account, levels of conceptual compatibility (or redundancy) reached 50% or more for all thesauri in the corpus. Although there is no benchmark figure available for this type of study, it seemed reasonable to see this as a particularly good level of compatibility in ILs of an encyclopaedic nature, and sufficient to conclude that conceptual overlap was high enough to justify the pursuit of research and development work on a common basic IL for non-art moving image collections. The fact that this thesaurus would be encyclopaedic in nature, and not domain restricted as is normally the case with this type of IL, would represent in itself a major challenge, but examples of such thesauri were becoming available at the time in the form of macrothesauri developed by national governments to provide access to their web sites (Hudon and Hjartarson 2002; Hudon 2005).

A third project allowed me to revisit issues of multilingualism and multiculturalism in ILs. Led by Professor Turner, the study consisted in comparing sets of index terms assigned independently by French speakers and by English speakers to the same set of images, and to compare these results to those obtained with automatic translation (Turner and Hudon 2002). A small scale experiment conducted in a controlled setting confirmed that the roughly 72% success rate obtained by the least efficient of the several translation software that were tested compared advantageously to the equivalence level of 69% obtained when comparing index terms provided by Francophone participants and equivalents provided by their Anglophone counterparts to describe the same image.

We suggested that automation of the index terms translation process in news agencies, television archives, and stockshot libraries with satisfying results was a clearly attainable goal, and established a link between this project and the previous ones; the potential application, in an automatic translation setting, of a general thesaurus to

represent the contents of non-art moving image collections. Such a thesaurus would raise even further the success rate obtained when automatically translating isolated names of objects, an operation far less complex than translation of full sentences. It would also contribute to better representation and retrieval of images on the web, using a diversity of natural languages.

## 6.0 Classification and the virtual library

From a knowledge and information organization perspective, the web constitutes an unbounded field of research and experimentation. For a period of 12 to 15 years around the turn of this century, subject directories and virtual libraries drew the attention of KO researchers looking for organization patterns in these new types of collections. In my own teaching and research, I used a simple definition of the concept of virtual library: collections of live links to digital resources that have been evaluated, selected, described, and classified by specialists (Hudon 2000, 2001b).

Although web organizers generally adopted the hierarchical model with remarkable enthusiasm, they did not believe that traditional classification methods and tools were essential or even appropriate to their needs. Category-based structures, the categories being disciplines or large topics, forms, audiences, names, etc., were preferred to discipline-based systems such as the *Dewey Decimal Classification* (DDC)<sup>3</sup> or the *Universal Decimal Classification* (UDC) schemes. Controlled inventories of standardized terms, such as list of subject headings or thesauri, were similarly ignored, superseded by keyword access to the full text of resources.

In a three-year project funded by the Fond québécois pour la recherche sur la société et la culture (FQRSC), I analyzed subject access structures in a sample of six virtual libraries in the field of Education, with the objective of detecting their strengths and weaknesses (Hudon 2003a, 2003b; Hudon, Mas and Gazo 2005; Hudon and Mas 2006). To do so, I applied a model developed by Sabine Mas, then doctoral candidate at Université de Montréal, as part of her work on the classification of electronic records residing on personal workstations in large organizations (Mas 2007; Mas and Hudon 2007). Strongly dependent on KO theory, the model allowed for a close examination of the classification schemes' structural, logical and semantic dimensions.

The structural analysis generated sets of quantitative data relating to maximum, minimum, and average numbers of classes and hierarchical levels. The numbers revealed little more than what was already known, or at least suspected. The average of 7.83 main classes in the six structures was well below the minimum of 10 top classes

deemed efficient for organizing resources in a specialized field, and not surprisingly, the higher numbers of distinct main classes were found in the most complete and complex classification structure. The average number of hierarchical levels, at 3.33, corresponded to a *de facto* standard number of levels recommended and common for general web classification structures; it is still assumed that a majority of information seekers will navigate to the third or fourth level of division only, each level corresponding to a mouse click, before reorienting their search. None of the six structures was overly complex and their lack of specificity did not allow for more than broad classification of the virtual collections.

Qualitative data relating to the logic of each structure was obtained through manual examination and interpretation by several coders working independently from each other. A mix of various principles of division (subject, object, audience, format (or external form), form of presentation, and location) was observed at the top three levels of all classification structures in the sample; the practice of mixing principles of division in a hierarchy is not recommended, because it generates classes that are not mutually exclusive, thus “causing uncertainty for the browser when he has to select a category” (Van der Walt 1998, 382). Objects, events, etc. were most frequently linked through relations of a contextual nature; this is also the case in traditional bibliographic classification schemes such as *DDC* and *UDC*. In a contextual relation, higher and lower classes are usually found within the same context but do not belong to the same taxonomy. This choice of a contextual, e.g. educational management > educational facilities, rather than a truly generic relation, e.g. educational institutions > secondary schools, contributes to making the structure more hospitable. The familiar alphabetical display of classes is also beneficial to hospitality, and undoubtedly preferable to a more or less obscure systematic arrangement reflecting the designer’s personal view of the world. Although the objectives of this project did not include that of comparing virtual and traditional structures on the basis of logic, my knowledge of the latter allowed me to suggest that the *ad hoc* structures were neither easier nor more difficult to navigate than *DDC* or *UDC*, which are considered complex and not user-friendly.

The semantic analysis provided data on conceptual and terminological concordance with authoritative sources in the field of education. Results were obtained through a standard methodology for estimating compatibility, involving manual examination of data and coder’s judgment as to degree of concordance. In this part of the analysis, because of the length and complexity of the task involved, only the most complex virtual classification structure was used that, of *The Educator’s Reference Desk (ERD)*. The structure was first aligned with the *DDC*, and then with the ta-

ble of contents of a trusted reference tool, the *Encyclopaedia of Educational Research*, 6<sup>th</sup> ed. Terminological and conceptual concordances were rather low, the latter being slightly greater, as expected. Partial concordance was greater than full concordance at both conceptual and terminological levels. Concordance between *ERD* and *DDC* was greater than concordance between *ERD* and the *Encyclopaedia*. The comparison with *DDC* obviously benefited from the universal character of the traditional scheme’s coverage; when a concept is only peripherally related to education, it may not be found in a specialized reference tool, but it is likely to be present somewhere in a general KO structure.

The virtual classification structures shared important weaknesses: a lack of concern for standardization, at the terminological and at the structural level; a lack of consistency in the arrangement of categories, subcategories, etc.; a disregard for theoretical principles, such as faceted analysis, citation order, mutual exclusivity, etc.; and, a lack of specificity, with users being led to long lists of unqualified and undifferentiated resources even at the lowest structural level, most likely a third level; the absence of relations between concepts.

This part of the project confirmed that the hierarchical model remained popular for organizing web resources in specialized virtual collections; hierarchies were contextual rather than generic, not overly complex and not very specific. Choice and organization of classes within the structures were systematic enough to make them easy to apprehend and navigate. But the structures were not very flexible and did not appear to benefit much from the technological environment in which they had developed and were now applied. Education level was the only principle of division that was used by all six virtual libraries. Two libraries only used explicit facets; they were however the least developed of all, and it was not possible to extrapolate on the usefulness of facets to describe and access the contents of virtual collections. A faceted model was explored in the second part of the project; it will be presented in section 8.0 of this paper.

## 7.0 Teaching classification

My teaching career spans four decades. During that time, the technology evolved at an incredible pace; powerful networks were created, leading to the implementation and expansion of the web. Librarians were replaced on the front lines by smart search engines, and users learned to access information from their living room, the neighbourhood café, or the train on their way to work. New types and formats of documents were offered, threatening the very existence of traditional collections. Through all the changes, LIS educators continued to advocate the impor-

tance of classification instruction, but it was obvious that classification could not anymore be presented mainly as a way to order physical collections, that it had to be promoted as an efficient means of providing subject access to all types of resources, including digital. Classification continues to be taught in LIS programs. If the subject matter itself has not changed much over the years, this is hardly the case for the context in which classification instruction is now offered. When I started teaching the principles of classification and the use of *DDC* and Library of Congress *Classification* (*LCC*), I was using a chalk and a blackboard, and my 20 students were poring over heavy books usually shared with several colleagues. Thirty-five years later, 120 students packed a computer lab, eyes glued to a screen, switching from one opened window to another, trying to figure out not only how to build *DDC* class numbers, but also the most efficient way to get to where they thought they ought to be in WebDewey®.

Classification calls for the application of three types of skills. On the cognitive level, students must learn how to analyze a subject, to identify concepts and facets, to discriminate between core and peripheral in relation to specific contexts, systems, and needs. On the technical level, students must learn to navigate classification structures and to translate concepts and subjects into appropriate representations in the form of class numbers or symbols. Technological skills are finally required of the students who learn to classify using web-based versions of classification schemes. In her research, Lisa Romero (1995) had observed that technical and technological skills were sufficiently developed in new graduates, but noticed a difficulty on their part to determine the subject of a document; her observations should have challenged all classification instructors.

In 2010, with the general objective of looking at the way bibliographic classification was now being taught, I launched a survey of classification instructors in ALA-accredited LIS masters' programs (Hudon 2011). A specific objective was to determine whether the cognitive process of subject analysis was given enough consideration in KO and classification courses. The survey was conducted via a web-based questionnaire structured around issues touching many dimensions of any teaching endeavour, with an emphasis on the manner in which students were trained in the development of the different types of skills required for classification. Course outlines were used as a secondary source of descriptive data. The analysis confirmed that: 1) classification retained its importance in LIS masters' programs where it was briefly introduced in core courses, and covered in depth in elective courses; 2) the amount of time dedicated to theory and principles remained modest, but these were indirectly presented in lessons on specific systems, with *DDC* and *LCC* used as pri-

mary examples of functional hierarchical structures; 3) instructors did not see the need to add theoretical content to their courses, but agreed on the necessity to integrate more examples and exercises; and, 4) the majority of instructors were confident that they were assessing correctly the subject analysis abilities of their students through their chosen evaluation methods.

In a literature review published at the start of this project (Hudon 2010), four areas in need of further investigation had been identified, the first of which was the goal of classification education. I agreed with several authors who suggested that educators had been so busy integrating new document types and formats, new technologies, new standards, etc. to the content of their classification courses, that they had lost sight of what it was exactly they were supposed to teach and of how and why they were supposed to do it. With this in mind, I set-up a follow-up study to examine general and specific objectives set by educators in classification courses. Its (Hudon and Guitard 2013, Hudon 2014) goal was to determine if and how the necessity for students to understand the basic principles and purposes of the classification process, and to acquire multiple sets of skills, was clearly reflected in course objectives.

Four hundred fifty-eight discrete objectives, gathered from 63 course outlines, were categorized along four dimensions: 1) nature (teaching objective / learning objective); 2) subject (general KO, descriptive cataloguing, access/subject access, classification/classification schemes, indexing/indexing languages, other); 3) focus (theoretical, analytical, technical/application, technological/interface, other); and 4) level, along the continuum proposed by Benjamin S. Bloom (1956). Verbs were used to determine the nature and taxonomic level of an objective, while nouns helped to identify subject and focus.

Four hundred objectives were coded as learning objectives; this large number is seen in a positive light since it indicates that classification instructors have adopted a learner- rather than a teacher-centered approach. A significant number of learning objectives were very general and rather vague, covering in a single sentence many concepts and subjects. In both core and elective courses, the proportion of objectives focusing on subject analysis (as a cognitive process) and on the use of technology was very low, as if instructors believed that these skills were acquired by instinct or osmosis rather than by formal instruction and dedicated practice.

Learning objectives may be ranked on the basis of the complexity of the cognitive processes required to achieve them. In this project, particular attention was given to the cognitive level of course objectives. Using a recent modification by Anderson and Krathwhol (2005) of the original Bloom's scale, each objective was coded to a level. The modified scale proposes six levels, from the least complex

to the most complex, from the concrete to the abstract; the levels are: remembering, understanding, applying, analyzing, evaluating, creating. It is expected that a majority of objectives set in higher education courses should sit at the higher and more abstract levels of the scale.

The analysis of the corpus revealed, however, that a large majority (88%) of learning objectives remained squarely in the lower levels of the scale, with as many as 308 objectives out of 400 written at level 2 and level 3; the verbs identify, understand, explain, describe, define, and apply were the most frequently used. This may seem adequate in core courses, but it is somewhat worrisome to observe that even elective courses, usually conceived as “advanced” courses, rarely venture beyond level 3 on the cognitive scale, the application level.

Given the limitations associated with the size of the corpus, no generalization has been possible. Tendencies, however, were revealed, and the results of the latest analysis reinforced observations made over the past 25 years in KO and classification education. While educators recognize and emphasize the necessity for students to develop high-level analytic and evaluative skills, explicit references to these skills in course objectives are few and far-between; this should provide all of us with food-for-thought.

### 8.0 Three encounters with facets

Contemporary networks constitute an ideal environment for implementing the analytico-synthetic principles and practices described by S.R. Ranganathan in the 1930s, as well as his proposal for a faceted classification structure. The digital environment is very flexible and it adapts easily to the evolution and renewal of concepts, subjects, and interests.

Like most KO researchers, I am attracted to the very notions of faceted analysis and facets, and I did explore the area several times over the past years. My first encounter with facets (Hudon and Mas 2001) happened at the end of the 1990s, in the framework of an applied research contract with the Quebec government, then in the process of re-engineering its records description and access systems with a view to integrating print and digital document. The overall goal of the re-engineering project was to recommend a more flexible way to access records besides that already provided by traditional archival classification based on provenance. Our small team of two was asked to validate the results of preliminary analyses made by internal working groups, to analyze existing records classification structures applied in government departments, to identify facets deemed essential for organizing and accessing government records, and to propose a methodology for the maintenance of controlled ILs.

Six classification schemes used within several government departments were first examined. There were surprisingly wide variations in the names of identical concepts within or among existing structures with little overlap. Concepts which should have been found in all existing structures, e.g. particular administrative activities such as planning and evaluation, were nowhere to be found in some schemes. These and other observations led to a recommendation that more control be exerted over the development and maintenance of a basic descriptive vocabulary.

Seven fundamental facets were identified: business process (e.g. budget, management), action (e.g., planning, control), object, concrete or abstract (e.g., vehicle, employee, document), document types (e.g., annual reports, minutes), agent (consumers, department), time/frequency and status/version (of a document). The first four were considered essential for describing and accessing government records. The facet agent was often implied, and thus not considered absolutely necessary. It was suggested that the last two facets be subsumed under document type.

My second encounter with facets occurred in the course of my study of the organization of virtual libraries in the field of education (described in Section 6.0 of this paper). The main objective of this second phase of the project was to propose a faceted structure to organize and access collections of web resources of interest to Francophone specialists and researchers in the field of education (Hudon 2007, 2008).

The faceted structure was developed in stages, as recommended among others by Vickery (1960) and Van der Walt (2004). A deductive approach dependent on literary warrant was chosen. A sample collection of 408 digital resources of various types (portals, journal articles, catalogues and bibliographies, institutional web sites, reference works, sponsored expert reports) was first created. Each resource was assigned to a DDC class and to a class in the hierarchical structured used in ERD, the most complete and complex structure in the virtual libraries that had been examined previously. The objective of the classification operation was to detect structural facets applicable to the field of education. Each resource was also assigned descriptors chosen in *EDUthès: Thésaurus de l'éducation* (<http://eduthes.cdc.qc.ca/>) a list of 4000 controlled terms describing the field. Once the tasks of classification and indexing had been completed, a reservoir of descriptors which could later be used as isolates in the faceted structure under development had been created.

The information acquired during previous stages of this research, as well as the set of titles and descriptors available in the sample collection, led to the identification of five fundamental facets considered essential to describe, structure and access a virtual collection of resources in the field

of education: agent (who?), activity (what?), method (how?), space (where?) and time (when?). To describe and access resources offering general information and knowledge, a sixth facet, foundations, was added. The structure was further developed using a division by essential characteristic rather than by contextual hierarchy. For example, teaching, an activity, was described from the perspectives of target clientele, objectives, educational level, subject taught, and teaching method. Class names were inspired by keywords extracted from titles and summaries in the sample collection, by *DDC* captions, by *ERD* class names, and by *EDUthès* descriptors. To preserve ease of access and user-friendliness, it was decided that the classification structure would not develop beyond the fifth level of subdivision, that the total number of isolates would not be higher than 400, and that the number of isolates at the most specific level of the structure would be as balanced as possible within each facet. Alphabetical order was preferred to systematic order of classes. The resulting faceted structure was constituted of six top-level facets representing as many departure points for navigating toward specific topics, 25 second-level classes, 86 third-level classes, 142 fourth-level classes and 62 classes at the fifth and lowest level of the structure.

The structure was then used to describe and organize all 408 documents in the sample collection. Although the lack of appropriate software and interface did not allow for optimal exploitation, the process confirmed that the structure could be evaluated positively in terms of potential extension and coverage, simplicity, navigational logic, and flexibility. A significant finding was that redundancy was an essential characteristic of such a structure; a concept, e.g. standards, would appear in several locations within the structure, possibly under different structural facets, with various but appropriate meanings and functions in context.

Although further testing of the faceted structure was obviously needed, it could not be carried as planned. The disappearance of most digital resources that had constituted our original sample, and the rapidly diminishing interest for virtual libraries themselves, made it difficult to justify the expense of acquiring software and deploying the means of testing and evaluation needed to go forward. It is remarkable that by the end of the second phase of the project, all virtual libraries in education examined two years earlier had either ceased to be maintained, or had been integrated to other, differently structured, web resources.

My third encounter (Hudon and Cumyn 2015) with facets is ongoing. It occurs in the framework of a five-year interdisciplinary team project funded by the Social Science and Humanities Research Council of Canada (SSHRC). The project is led by Université Laval, Québec, law Professor Michelle Cumyn, who has developed a strong interest for the difficulties linked to information research in con-

temporary legal databases. Law is a complex domain of knowledge, the structure and strict organization of which are not immediately accessible to non-specialists, such as members of the general public and even law students. Systematic organization of legal information sources is partly based on legal regimes (law of obligations, contract, sale, guarantee of quality, etc.), considered essential in the search for sources relevant to specific legal problems. Categorization also relies on the representation of facts in precedents and by authority. Regimes and precedents are grouped within more general categories corresponding to traditional fields and types of law, such as civil or criminal.

In law as in other domains, the creation of digital libraries and the possibility for users to access the full text of documents has led to the erroneous belief that complex classification schemes were not needed anymore. Producers provide access to legal databases through categories corresponding, more or less, to fields of legal practice. This choice is justified, on the one hand, by the non-availability of a classification structure or taxonomy recognized by all who study and practice law and, on the other hand, by the perceived reluctance of users to navigate a complex organization structure. When in need of legal information, individuals tend to rely on keywords extracted from the description of a situation, e.g. fall sidewalk ice, rather than drilling down through a hierarchy of legal categories (municipal law, civil responsibility, etc.).

Broughton (2010) believes that law, as a discipline, is particularly well-suited to faceted organization; the number of facets required to structure legal information is most likely limited. The ultimate aim of the project is to propose a new system, structured around facets, to describe and access the content of legal information sources. We hypothesize that such a system has the potential to be more user-friendly and efficient than existing ones. A secondary objective is to make it easier for non-specialists to link descriptions of facts or situations generating legal problems to applicable legal regime(s), thus making it easier to access and navigate the complex structure of the law.

A year into the project, the facets that will be tested are emerging from a thorough review of literature and research, combined to an analysis of existing ILs and legal reference tools. It is already clear that two sets of facets will be needed. A first set of three or four facets will be used to describe the functional components of the fact or situation generating the legal problem. The second set of three or four facets will link to the structure of the law. Factual and legal facets will interact dynamically, as the information search progresses through choice and combination of specific isolates.

Members of the team believe that the faceted organization will allow users to better structure and focus their search strategies. The findings should be of interest to

law teachers, since the faceted organization would provide students with a connection between a common, familiar representation of a legal problem or situation, and the formal, complex structure of the legal system they must learn to navigate. Furthermore, the use of facets could make legal information databases destined to the general public more attractive and accessible. Facets should facilitate multi-database searching and interoperability, even where legal systems and regimes differ.

This interdisciplinary project will contribute to the ongoing reflexion on the usefulness, usability, and efficiency of faceted analysis and faceted structures. It is also a great opportunity to explore a domain of knowledge which has not been studied in depth by the KO community, and which certainly presents interesting challenges.

## 9.0 To conclude ... for now

I'll admit to being worried at first that this forced backward look at the past decades would result in dissatisfaction with what I had done in that period of time, both in terms of quantity and quality of work. I built my career around research topics and projects that were feasible, that attracted me, and that would bring pleasure and satisfaction, rather than around projects that absolutely needed to be conducted, were guaranteed to attract funding or would generate a maximum of citations.

As a result, my body of work will be judged unremarkable by any standard and my contribution to the advancement of the field of knowledge organization not as significant as I would have liked it to be. But I am particularly proud of my work on multilingual/multicultural ILS which inspired a few research projects and applications, and of my survey of classification instruction in LIS programs; the latter has provided data that could be used as a departure point in much-needed further studies and recommendations relating to this all important curriculum subject and professional practice.

The selection of projects described here, others on which I have worked over the years, and my whole career as professional librarian, teacher, and researcher, all bear witness to my belief that relevant information can only become truly accessible when it has been described, indexed, and classified, either by a human analyst or by an automated system that has learned the rules and can make "intelligent" use of authority files in some form or other.

Before concluding, I must acknowledge a few individuals who played most important roles in this unintended but ultimately gratifying career. Professor Suzanne Bertrand-Gastaldi introduced me to the thesaurus, well before it became as popular as it would be in the 1980s. Professors Nancy Williamson and Clare Beghtol guided my first steps on the path of scientific research. Professor Elaine

Svenonius, a most gracious examiner of my dissertation, was instrumental in my decision to seek an academic position after my doctoral work instead of going back to the world of thesaurus construction and management. The influence of Jacques Maniez's work on my own has been most important; he was much amused when I called him my "hero" at an ISKO-France colloquium; he became a friend and I will always treasure our animated discussions of indexing, indexing languages, and facets.

My teaching career has now ended, but my work as researcher will hopefully continue for a little while. The world of indexing languages and subject representation offers an infinite number and variety of research topics, many of which I still want to read and write about. And I hope to see more young researchers choose this rich field of investigation.

## Notes

1. The Cégep is a post-secondary institution exclusive to the education system in Québec. Pre-university programs are two years in duration. Technical programs are three years in duration, with specializations leading to a career right after graduation.
2. This work was carried out under a Steven I. Goldspiel Memorial Research Grant, awarded in 1999 by the Special Libraries Association, Washington, DC.
3. Dewey, *Dewey Decimal Classification*, DDC, OCLC and WebDewey are registered trademarks of OCLC.

## References

- Bloom, Benjamin S. 1956. *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain*. New York: David McKay.
- Broughton, Vanda. 2010. "The Use and Construction of Thesauri for Legal Documentation." *Legal Information Management* 10:35-42.
- Hooper, R. S. 1966. "Evaluation and Analysis of Indexing Systems." In *The Second Institute on Technical Literature Indexing, Session 1*. Washington, D.C.: American University, Center for Technology and Administration, 16-18.
- Hudon, Michèle. 1996. "Preparing Terminological Definitions for Indexing and Retrieval Thesauri: A Model." In *Knowledge Organization and Change: Proceedings of the Fourth International ISKO Conference, 15-18 July 1996, Washington, DC.*, ed. R. Green. Advances in knowledge organization 5. Frankfurt/Main: Indeks, 363-9.
- Hudon, Michèle. 1997. "Multilingual Thesaurus Construction: Integrating the Views of Different Cultures in One Gateway to Knowledge and Concepts." *Knowledge Organization* 24:84-91.

- Hudon, Michèle. 1998a. "A Preliminary Investigation Of the Usefulness of Semantic Relations and of Standardized Definitions for the Purpose of Specifying Meaning in a Thesaurus." In *Structures and Relations in Knowledge Organization: Proceedings of the Fifth International ISKO Conference, 25-29 August 1998, Lille, France*, ed. W. Mustafa El Hadi, W. J. Maniez and A. S. Pollitt. Advances in knowledge organization 6. Würzburg: Ergon, 139-45.
- Hudon, Michèle. 1998b. "Compatibility is not a Synonym of Identity: Conceptual Structures in Multilingual Thesauri." *Knowledge Organization* 25:152-5.
- Hudon, Michèle. 1999. "Accessing Documents and Information in a World without Frontiers." *The Indexer: The International Journal of Indexing* 21:156-9.
- Hudon, Michèle. 2000. "Innovation and Tradition in Knowledge Organization Schemes on the Internet, or, Finding One's Way in the Virtual Library." In *Dynamism and Stability in Knowledge Organization: Proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*, ed. C. Beghtol, L. Howarth and N. J. Williamson. Advances in knowledge organization 7. Würzburg: Ergon, 35-40.
- Hudon, Michèle. 2001a. "Semantic Relationships in Multilingual Thesauri." In *Relationships in the Organization of Knowledge*, ed. C. Bean and R. Green. Dordrecht, The Netherlands: Kluwer Academic Publishers, 67-80.
- Hudon, Michèle. 2001b. "Structuration des savoirs et organisation des collections dans les répertoires du Web." *Bulletin des bibliothèques de France* 46, 1:57-62.
- Hudon, Michèle. 2003a. "Subject Access to Web Resources in the Field of Education." In *Subject Retrieval in a Networked Environment: Proceedings of the IFLA Satellite Meeting, Dublin, OH, 14-16 August 2001*. München, Germany: K.G. Saur, 83-9.
- Hudon, Michèle. 2003b. "Expanding Audiences for Education-Related Information and Resources: Classificatory Structures and Education Thesauri on the World Wide Web." *Cataloging and Classification Quarterly* 37, nos. 1-2:81-96. [http://dx.doi.org/10.1300/J104v37n01\\_07](http://dx.doi.org/10.1300/J104v37n01_07).
- Hudon, Michèle. 2004. "Conceptual and Lexical Compatibility in Thesauri Used to Describe and Access Moving Image Collections." In *Access to Information: Technologies, Skills, and Socio-Political Context: Proceedings of the 32nd Annual Conference of the Canadian Association for Information Science, 3-5 June 2004, Winnipeg, Manitoba*. <http://www.caais-acs.ca/ojs/index.php/cais/article/view/11/626>
- Hudon, Michèle. 2005. "Macrothésauri et systèmes d'information gouvernementaux accessibles au grand public." In *L'organisation des connaissances : Approches conceptuelles : Actes du 4<sup>e</sup> congrès de ISKO-France, 3-4 juillet 2003*. Paris: L'Harmattan, 213-22.
- Hudon, Michèle. 2006. "Le thésaurus, au carrefour des sciences de l'information et de la terminologie." In *Terminologie et accès à l'information spécialisée*. Sous la dir. de W. Mustafa El-Hadi. Paris: Hermès Science, 71-98.
- Hudon, Michèle. 2007. "Proposal for an Alternate Structure to Organize and Access Resources in a Virtual Library in Education." In *Information Sharing in a Fragmented World: Crossing Boundaries: Proceedings of the 35<sup>th</sup> Annual Conference of the Canadian Association for Information Science, 10-12 May, 2007, Montréal, QC*. <http://www.caais-acsi.ca/ojs/index.php/cais/article/view/695/452>
- Hudon, Michèle. 2008. "Using Facets to Classify and Access Digital Resources: Proposal and Example." In *Digital Libraries*, ed. Fabrice Papy. London, England: ISTE / Wiley, 145-67.
- Hudon, Michèle. 2010. "Teaching Classification, 1990-2010." *Cataloging & Classification Quarterly* 48, no. 1:64-82. <http://dx.doi.org/10.1080/01639370903356370>
- Hudon, Michèle. 2011. "Teaching Classification in the 21<sup>st</sup> Century." *Knowledge Organization* 38: 342-51.
- Hudon, Michèle. 2012. "ISO25964: pour le développement, la gestion et l'interopérabilité des langages documentaires." *Documentation et bibliothèques* 58:130-40.
- Hudon, Michèle. 2014. "KO and Classification Education in the Light of Benjamin Bloom's Taxonomy of Learning Objectives." In *Knowledge Organization in the 21<sup>st</sup> Century: Between Historical Patterns and Future Prospects: Proceedings of the 13<sup>th</sup> Conference of the International Society for Knowledge Organization (ISKO 2014), Krakow, Poland, 19-22 May 2014*, ed. W. Babik, Advances in knowledge organization 14. Würzburg: Ergon Verlag, 524-31.
- Hudon, Michèle and Michelle Cumyn. 2015. "Organisation et accès à l'information juridique à l'ère du numérique." In *Systèmes d'organisation des connaissances et humanités numériques : Actes du 10<sup>e</sup> Colloque ISKO-France*, 5-6 novembre 2015, Strasbourg, France. [In press]
- Hudon, Michèle and Laure Amélie Guitard. 2013. "KO and Classification Instruction Objectives: Are We Keeping Up with the Transformation of Our Field?" In *Transition Cultures, Transition KO: Evolving Exploration, Critical Reflection, and Practical Work: Proceedings of the 4<sup>th</sup> NASKO (North American Chapter of ISKO) Symposium on Knowledge Organization, Milwaukee, WI, 13-14 June 2013* ed. Richard P. Smiraglia. <http://dx.doi.org/10.7152/nasko.v4i1.14651>
- Hudon, Michèle and Fay Hjartarson. 2002. "Governments Meet People: Developing Metathesauri in the Framework of Government Online Initiatives." In *Advancing Knowledge: Expanding Horizons for Information Science: Proceedings of the 30th Annual Conference of the Canadian Association for Information Science, 30 May - 1 June 2002, Toronto*,

- Canada. Toronto, ON: CAIS, 46-60. <http://www.cais-acsi.ca/ojs/index.php/cais/article/view/479>
- Hudon, Michèle and Sabine Mas. 2001. *Analyse des facettes pour la classification des documents institutionnels au gouvernement du Québec: Rapport préparé pour le Groupe de travail en classification et indexation*. Québec. Secrétariat du Conseil du trésor, Sous-secretariat à l'inforoute gouvernementale et aux ressources informationnelles.
- Hudon, Michèle and Sabine Mas. 2006. "Structure, Logic, and Semantics for Web-Based Collections in Education." In *Knowledge organization for a global learning society: Proceedings of the Ninth International Society for Knowledge Organization (ISKO), 4-6 July 2006, Vienna, Austria*, ed. G. Budin, C. Swertz and K. Mitgutsch. Advances in knowledge organization 10. Würzburg: Ergon, 263-70.
- Hudon, Michèle, Sabine Mas and Dominique Gazo. 2005. "Structure, Logic, and Semantics in *Ad Hoc* Classification Schemes Applied to Web-Based Libraries in the Field of Education." *Canadian Journal of Information and Library Sciences / Revue canadienne des sciences de l'information et de bibliothéconomie*, 29:265-88.
- Hudon, Michèle, J. M. Turner and Yves Devin. 2000. "How Many Terms Are Enough? Stability and Dynamism in Vocabulary Management for Moving Image Collections." In *Dynamism and Stability in Knowledge Organization: Proceedings of the Sixth International ISKO Conference, 10-13 July 2000, Toronto, Canada*, ed. C. Beghtol, C. L. Howarth and N. J. Williamson. Advances in knowledge organization 7. Würzburg: Ergon, 333-8.
- Hudon, Michèle, J. M. Turner and Yves Devin. 2001. "Description et indexation des collections d'images en mouvement: résultats d'une enquête." *Documentation et bibliothèques* 47: 5-12.
- Krathwohl, David R. 2005. "A Revision of Bloom's Taxonomy: An Overview." *Theory into Practice* 41:212-8.
- Mas, Sabine. 2007. *Schémas de classification et repérage des documents administratifs électroniques dans un contexte de gestion décentralisée des ressources informationnelles*. Thèse de doctorat. Montréal: Université de Montréal, École de bibliothéconomie et des sciences de l'information. <http://hdl.handle.net/1866/1433>
- Mas, Sabine and Michèle Hudon. 2007. "Schémas de classification et repérage de documents électroniques: pratiques individuelles et application des principes classificatoires archivistiques." In *Information Sharing in a Fragmented World: Crossing Boundaries: Proceedings of the 35<sup>th</sup> Annual Conference of the Canadian Association for Information Science, 10-12 May, 2007, Montréal, QC*. <http://www.cais-acsi.ca/ojs/index.php/cais/article/view/706>
- Riesthuis, G. J. H. 1996. "Theory of Compatibility of Information Languages." In *Compatibility and Integration of Order Systems: Research Seminar Proceedings of the TIP/ISKO Meeting, Warsaw, 13-15 September, 1995*. Warsaw: WYDAWNICTWO SBP, 23-31.
- Romero, Lisa. 1995. "An Evaluation of Classification and Subject Cataloging in Entry-Level Cataloging Copy: Implications for Access and Instruction." *Journal of Education for Library and Information Science* 36:217-29.
- Sager, J. C. 1982. "Terminological Thesaurus: A More Appropriate Designation or a Deprecated Synonym?" *Social Science Information Studies* 2:211-4.
- Sager, J. C. 1990. *A Practical Course in Terminology Processing*. Amsterdam, The Netherlands: John Benjamins.
- Sager, J. C. and M. C. L'Homme. 1994. "A Model for the Definition of Concepts: Rules for Analytical Definitions in Terminological Databases." *Terminology* 1:351-73.
- Turner, James M. and Michèle Hudon. 2002. "Multilingual Metadata for Moving Image Databases." In *Advancing Knowledge: Expanding Horizons for Information Science: Proceedings of the 30th Annual Conference of the Canadian Association for Information Science, 29 May - 1 June 2002, Toronto, Canada*. Toronto, ON: CAIS, 34-45.
- Van der Walt, M. S. 1998. "The Structure of Classification Schemes Used in Internet Search Engines." In *Structures and Relations in Knowledge Organization: Proceedings of the Fifth International ISKO Conference, 25-29 August 1998, Lille, France*, ed. W. Mustafa El Hadi, W. J. Maniez and A. S. Pollitt. Advances in knowledge organization 6. Würzburg: Ergon, 379-87.
- Van Der Walt, M. S. 2004. "A Classification Scheme for the Organization of Electronic Documents in Small, Medium, and Micro Enterprises (SMMEs)." *Knowledge Organization* 31:26-38.
- Vickery, Brian Campbell. 1960. *Faceted Classification: A Guide to Construction and Use of Special Schemes*. London: ASLIB.