A Second Life for TIIARA: From Bilingual to Multilingual!

Elaine Ménard*, Nouf Khashman**, Svetlana Kochkina***,
Juan-Manuel Torres-Moreno****, Patricia Velazquez-Morales*****, Fen Zhou******,
Pierre Jourlin*******, Priyanka Rawat********, Peter Peinl*********,
Elvys Linhares Pontes**********, Ilaria Brunetti***********

* McGill University, School of Information Studies, Canada, <elaine.menard@mcgill.ca>
**Qatar National Library, Qatar Foundation, <nkhashman@qf.org.qa>
*** McGill University, Gelber Law Library, Canada, <Svetlana.kochkina@mcgill.ca>
****Laboratoire Informatique d'Avignon, UAPV, France and École Polytechnique de Montréal, Canada, <juan-manuel.torres@univ-avignon.fr>
*****VM Labs, France, <patricia_velazquez@yahoo.com>
******Laboratoire Informatique d'Avignon, UAPV, France, <fen.zhou@univ-avignon.fr>
*******Laboratoire Informatique d'Avignon, UAPV, France, <pierre.jourlin@univ-avignon.fr>
********Laboratoire Informatique d'Avignon, UAPV, France, <priyankaarawat@gmail.com>
********* University of Applied Sciences, Hochschule Fulda, Deutschland, <Peter.L.Peinl@informatik.hs-fulda.de>
**********Laboratoire Informatique d'Avignon, UAPV, France, <elvys.linhares-pontes@alumni.univ-avignon.fr>
***********Laboratoire Informatique d'Avignon and INRIA, France, <ilaria.brunetti@inria.fr>

Elaine Ménard is Associate Professor at the School of Information Studies, McGill University, Montreal, Canada. Her teaching expertise includes cataloguing, indexing, classification and information retrieval. Her main research interests deal with cross-language information retrieval, image indexing and metadata.

Nouf Khashman is Senior Information Services Librarian for Learning Resources at Qatar National Library (QNL), responsible for developing and coordinating programs and services for the community, and leading the activities of the Learning Commons and the Adaptive Technology and Services. Prior to joining QNL, Nouf worked on several research projects related to human-information interaction, including information visualization and bilingual information retrieval. Her research interests lay in design, user experience, and usability (DUXU) and cross-cultural communication.

Svetlana Kochkina is a PhD candidate at the McGill School of Information Studies and a librarian at the McGill University Law Library. Her main research interests include: evolution of the form and paratextual elements of the book; history of publishing, reading, and book collecting in cultural and social context; history of institutional and private libraries.

Juan-Manuel Torres-Moreno, PhD in computer science at Institute National Polytechnique de Grenoble, Head of Research group of TALNE (Laboratoire Informatique d'Avignon) of the Université d'Avignon, France and specialist in automatic text summarization, automatic text classification, information retrieval and machine learning.

Patricia Velazquez-Morales, PhD in materials science from the Institut National Polytechnique de Grenoble, and specialist in automatic text summarization and creation of specialized multilingual corpus.

Fen Zhou received a PhD degree in computer science from INSA Rennes (France) in 2010. He is currently an associate professor at the Université d'Avignon, France. His research interests include combinatorial optimization problems in computer networks and intelligent transportation systems.

Pierre Jourlin obtained a PhD in computer science in 1998 in the domain of speech recognition. He worked as a researcher at the University of Cambridge (UK) on information retrieval from 1997 to 2000 and has been a lecturer in the Université d'Avignon since 2000. His current area of research is text mining.

Priyanka Rawat received a B.Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, India and a PhD degree in computer science from Telecom Bretagne, France. She is currently an assistant professor at the Université d'Avignon. Her research interests include wireless and sensor networks, cognitive radio for vehicular ad hoc networks, and heterogeneous networks.

Peter Peinl received a PhD in computer science in 1986 from Technical University Kaiserslautern, Germany. He became a full professor of database systems and information retrieval at Fulda University of Applied Science, Fulda, Germany in 1996. His research interests include database systems and information retrieval.

Elvys Linhares Pontes received a master's degree in computer engineer from the Federal University of Ceará (Brazil) in 2015. In the current moment, he is a PhD student at the Université d'Avignon (France) in Automatic Text Summarization and Information Retrieval.

Ilaria Brunetti, PhD in computer science at Université d'Avignon and INRIA Sophia Antipolis in 2015. Her research interests include game theory, in particular evolutionary game theory and decision dynamics.

Ménard, Elaine, Khashman, Nouf, Kochkina, Svetlana, Torres-Moreno, Juan-Manuel, Velazquez-Morales, Patricia, Zhou, Fen Jourlin, Pierre, Rawat, Priyanka, Peinl, Peter, Linhares Pontes, Elvys, and Brunetti, Ilaria. "A Second Life for TIIARA: From Bilingual to Multilingual!" *Knowledge Organization Knowledge Organization* 43 no. 1: 22-34. 39 references.

Abstract: Multilingual controlled vocabularies are rare and often very limited in the choice of languages offered. TIIARA (Taxonomy for Image Indexing and RetrievAl) is a bilingual taxonomy developed for image indexing and retrieval. This controlled vocabulary offers indexers and image searchers innovative and coherent access points for ordinary images. The preliminary steps of the elaboration of the bilingual structure are presented. For its initial development, TIIARA included only two languages, French and English. As a logical follow-up, TIIARA was translated into eight languages—Arabic, Spanish, Brazilian Portuguese, Mandarin Chinese, Italian, German, Hindi and Russian—in order to increase its international scope. This paper briefly describes the different stages of the development of the bilingual structure. The processes used in the translations are subsequently presented, as well as the main difficulties encountered by the translators. Adding more languages in TIIARA constitutes an added value for a controlled vocabulary meant to be used by image searchers, who are often limited by their lack of knowledge of multiple languages.

Received: 14 September 2015; Revised: 30 November 2015; Accepted 1 December 2015-12-10

Keywords: language, languages, translation, TIIARA, taxonomy, multilingual

1.0 Introduction

With the growing amount of non-textual objects produced, both personal and public, the organization and retrieval of these items has become imperative. Non-textual objects can include photography, video recordings, traditional visual art (paintings, sculptures, installations), cultural artifacts, music and so on. The organization of visual resources such as personal photos could be problematic. Traditional controlled vocabularies have been used for indexing and classification of visual information such as images for years, with a variable degree of precision. They are mainly used to identify a primary subject or the broader category of an item, that is, critical metadata associated with any image being described. Among established vocabularies used for image indexing, there are the *Library of Congress Subject Headings* (Library of Congress 2015a), Getty's *Art & Architecture Thesaurus*® *Online* (Getty Research Institute 2015a) and the *Thesaurus for Graphic Materials* (Library of Congress 2015b). Although very well structured and developed, these terminologies are not always accurate for the very specific type of images that are "ordinary images," such as the pictures accumulated over the years with digital cameras. Furthermore, if a variety of thesauri, subject heading lists and authority lists exists, there is no guarantee these tools will provide terms in many languages. In reality, most controlled vocabularies used for indexing textual and non-textual documents offer a terminology in one language only, with some exceptions. For example, the *UNESCO Thesaurus* is a multilingual (English, French, Spanish and Russian) (UNESCO 2015) controlled and structured list of terms used in subject analysis and retrieval of documents and publications in the fields of education, culture, natural sciences and social and human sciences. It is also worth mentioning the Getty's *Art & Architecture Thesaurus*® *Online* (Getty Research Institute 2015a) contains contributions in Spanish from Centro de Documentación de Bienes Patrimoniales, Chile; Dutch from the Rijksbureau voor Kunsthistorische Documentatie; a Chinese translation is underway by the National Digital Archives Program, Taiwan; German translation is being undertaken by the Institut für Museumsforschung in Berlin; around 3,000 Italian object type terms were contributed by ICCD, Rome; and 3,000 French terms were contributed by the Canadian Heritage Information Network. However, authentic multilingual vocabularies are rare and often very limited in the choice of languages offered.[1] Therefore, these linguistic restrictions will have obvious consequences on their use for indexing and retrieval worldwide.

In order to fill this double gap (Ménard 2012), TIIARA (Taxonomy for Image Indexing and RetrievAl), a bilingual taxonomy dedicated to image indexing was developed. The main goal of this project was to create a vocabulary that would offer indexers and image searchers innovative and coherent access points for ordinary images. Indeed, this taxonomy could be a useful feature for searching images. Very few search engines offer their users the opportunity to browse a taxonomic structure to initiate or refine queries. This desire to search for images with a predetermined list of subjects (taxonomy) was expressed by many image searchers (Ménard et al. 2013) who participated in the exploration of the roles and usefulness of functionalities for image retrieval in a bilingual context. The main advantages (Jörgensen 2003) of controlled vocabularies such as a taxonomy are to promote consistency and to increase the probability of matching words chosen by the indexer to those of the image searcher. Inevitably, using a taxonomy for the image retrieval process will facilitate the searching process.

For its initial development, TIIARA included only two languages, French and English. Even if language differences often also imply cultural and conceptual differences, it was not necessarily the case with these two languages, which are the two official languages of Canada, where this research is taking place. However, problems with the structural hierarchy could arise in multilingual controlled vocabularies, particularly when the different languages included show crucial variations in their hierarchical levels where concepts are organized. Nevertheless, it was decided, as a logical follow-up, to continue the TIIARA enhancement and integrate more languages to increase its international scope.

This paper describes the different stages of this development. In the first section, the preliminary elaboration of the bilingual structure is described. In the following section, the steps taken to test and update TIIARA are presented. In section 4, the different processes used in the translations are exposed. The main difficulties encountered during the translations are discussed in section 5, and section 6 concludes the paper and proposes future directions to improve the multilingual taxonomy.

2.0 The "making of" the bilingual taxonomy

The first structuring of the taxonomy involved the choices of top-level categories and their subcategories. Two approaches are usually considered: the bottom-up approach and the top-down approach. The first strategy supposes starting from the narrowest terms possible and moving to the more generic ones, while in the second method, the concepts are named and placed within the taxonomy, as part of its organization, and then subdivided. With the top-down approach, the attempt is made to first represent some of the main concepts and then identify narrower terms to eventually reach the desired level of granularity. In other words, the necessary hierarchical structure—as well as the relationships—is created as the development of the vocabulary takes place.

For the development of TIIARA, both approaches were combined. The starting point (Ménard and Smithglass 2014) was a list of potential concepts resulting from a broad analysis of existing specialized terminologies used by professional indexers to describe images, as well as tags employed by regular Internet users. The appropriate categories were extracted and the values were clustered into a limited number of groupings. This shaped the basis from which categories, subcategories and candidate descriptors were then selected. The following were included from the listing of concepts: taxonomy perspectives (top-down); shared characteristics of concepts, particularly those from users or content (bottom-up); and high-level unification (top-down). These combined approaches identified gaps that needed to be filled, the hierarchy and the alternative entry points. As a result, this first version of TIIARA offered a coherent structure produced simultaneously with the choice and definition of top-level categories and their subcategories.

Rather than developing the taxonomy in one language and having it translated once completed, it was decided that the selected categories and subcategories would be developed simultaneously in English and French. In other words, the bilingual structure of the taxonomy was kept as parallel as possible. This strategy seemed manageable (Walter 2001), since both English and French come from the same Indo-European language family and have common origins. The Indo-European family includes languages largely used throughout Europe and other parts of the world as a result of colonization. This group of languages (Ethnologue 2015) refers to the easternmost extension of the family from the Indian subcontinent to its westernmost reach in Europe. Among the many branches of this linguistic family, it is worth mentioning the Romance languages that were shaped from Vulgar Latin dialects spoken by the common people in the Roman Empire. This branch includes five widely spoken languages (Ethnologue 2015) (by number of native speakers): Spanish (410 million), Portuguese (216 million), French (75 million), Italian (60 million) and Romanian (25 million). The Indo-European family also comprises the Germanic language "branch" that counts more than 470 million speakers in many parts of the world but mainly in Europe and the Western Hemisphere. The English language belongs to that division of the Indo-European language family. It is most closely related to the German and Dutch languages.

Coming from the same large family, English and French are not that different theoretically. However, dialectal and cultural dimensions needed to be considered in the development of TIIARA. For example, the level of language (e.g., Canada/United States versus United Kingdom; Quebec versus France) may not be considered universal. Indeed, even if the same concepts and messages are still expressed in the same way, differences exist inside a same language and localization could be needed. Localization (Jiménez-Crespo 2013) is the process used to adjust a product or service to a particular context. In localizing a product, in addition to idiomatic language translation, details such as time zones, currencies, national holidays, local colour sensitivities, product or service names, gender roles and geographic examples all must be considered.

After the first structuring phase, the taxonomy included essential facets that appropriately represent ordinary images in order to increase the effectiveness and the efficiency of image browsing and searching. Finally, definitions of the concepts included in the taxonomy were provided in order to enhance its understanding. Nevertheless, since the development of a taxonomy, like any other form of controlled vocabulary, is an iterative process. Incremental user testing needed to be carried out in order to validate and refine the categories, values and relationships. TIIARA was validated using a number of different techniques, samples of images and users (indexers and image searchers). The main objective of these validating processes was to ensure that the taxonomy was clear, easy to use and consistent.

3.0 Testing and updating

The initial version of TIIARA included nine main categories and went through eight revisions before it was considered stable enough to be tested. At this stage, the bilingual taxonomy was evaluated with 12 participants using the card-sorting technique. The sample included six French-speaking and six English-speaking participants, six females and six males, between the ages of 19 and 30. For this experimentation, the nine first-level categories were written on labels and placed on a table and all participants were instructed to place all second-level subcategories under one of the provided categories by the criteria of their choice. All of the subcategories had to be placed under one of the provided categories, although it was not mandatory that all first-level categories be used. This testing was expected to provide a basis for comparison between relationships made by the participants versus those that were made with our predetermined structure. After the sorting, respondents were asked to comment on their experience and make suggestions for improving the taxonomy.

The analysis of the sorting, as well as the comments and suggestions received after the card-sorting exercise, highlighted several elements that needed to be taken into consideration and a first round of modifications was completed. The preliminary nine main groupings were reduced to seven, as indicated in Table 1.

Main Category	Definition
Abstract Ideas Idées abstraites	Related to an idea of something formed by mentally combining all its characteristics or particulars; a concept.
Arts and Entertainment Art et divertissement	Related to people, tools, equipment and products specifically associated with dance, design, visual arts, writing, music, television and film, and stage.
Daily Life Vie quotidienne	Related to the activities and experiences that constitute a person's normal existence.
Nature Nature	Related to the phenomena of the physical world, including plants, animals, the landscape and other features and products of the Earth, as opposed to humans or human creations.
Places Lieu	Related to a building or a physical environment used for a special purpose.
Objects and Equipment Objet et équipement	Related to unique objects or pieces of equipment (not in active use by a person).
Work Travail	Related to people doing a job, other than those listed in "Arts and Entertainment."

Table 1. TIIARA main categories and definitions

These seven main categories were then developed to include second-, third- and, in some cases, fourth-level subcategories. It was decided to keep the number of top-level categories and the depth of the taxonomic structure to a minimum in order to avoid potential frustration for the eventual users. This refinement of the categories was conducted simultaneously in English and French. At this stage, TIIARA included essential facets that appropriately described ordinary images. The subcategories also offered the appropriate level of granularity in order to improve user searches.[2]

The next phase involved the indexing of a small image database (Image Database Donated Liberally [IDOL], which includes 6,015 ordinary images) by two indexers (one English and one French native speaker). A detailed comparison of the indexing terms assigned by the indexers was then undertaken. Although this analysis revealed potential holes in the taxonomy (Ménard 2013), it also highlighted that TIIARA was already providing sufficient vocabulary to index images at the appropriate level of specificity. Subsequently, TIIARA was retested with a sample of 60 respondents (30 English-speaking and 30 French-speaking). The participants were shown 30 images randomly selected from the database in the same order of presentation and were asked to indicate where in the taxonomic structure they thought they would find each one of the images. Once the retrieval simulation was completed, participants answered a questionnaire to give their general opinion on TIIARA. The questionnaire included 12 closed questions with responses indicated on Likert scales and four open-ended questions that asked users to provide feedback about TIIARA. The questionnaire evaluated the quality of the entire taxonomy as well as the overall satisfaction from an image searcher's perspective.

The data collected in this phase of TIIARA testing (Ménard and Dorey 2014) revealed that all of the 30 images were correctly retrieved by at least one English-speaking participant and 27 out of the 30 images were correctly retrieved by at least one French-speaking participant. There were no differences in the average amount of time to correctly retrieve images between the English and French groups, with both groups taking on average 19 seconds. English-speaking participants on average took nearly 24 seconds per attempt for those cases when the image was not correctly retrieved, and French-speaking participants on average took 23 seconds. On average, English-speaking participants took 1.58 attempts to correctly retrieve an image, and French-speaking participants took 1.60 attempts.

Following the analysis of the data collected during that second phase of testing, TIIARA was updated using not only the findings of the retrieval performance but also the feedback received from the indexers that described the 6,015 images of the collection using TIIARA. Their points of view were essential in order to improve the taxonomy. Indeed, during the indexing process, some gaps were revealed. The first clue was the proportion of images that received an indexing term from a very general category. This exposed where subdevelopment was needed. Moreover, sometimes the categories were not granular enough and revealed missing terms. For example, the indexers used the term "Libraries" from the "Work" main category for a library building, even if the picture did not include any library worker as defined by the scope note. In the same way, they used "Agriculture" for pictures of farms or fields of crops even though there were no agricultural workers in the picture. These types of holes needed to be filled; otherwise, the image searchers who would browse the taxonomy would never be able to find the picture of a library or a farm in the main category "Places," where it needed to be included. In other words, the image searcher would probably not have the reflex to browse the main category "Work" that is "related to people doing a job" to find the picture of a library or a farm. Finally, indexers also felt that some subcategories were underdeveloped. For example "Visual Artworks" could include more subcategories in order not to become the last resort for all pictures of art.

In addition to the indexers' reactions provided following the indexing process, other oversights have been addressed (Ménard and Girouard 2015) thanks to the comments received from the people who participated in the TIIARA testing. It is worth mentioning that, on the one hand, the taxonomic structure itself was not modified, that is, the hierarchical levels remained the same. On the other hand, some subcategories at the third or fourth levels were added to provide better specificity. For example, many terms were integrated in the most-used categories ("Places," "Daily Life" and "Nature"). However, it is fully understood that the processes of updates and maintenance (Aitchison et al. 2000), as it is for most controlled vocabularies, have to be continuous to ensure that the terminology always remains current and inclusive.

Once TIIARA was updated and retested, it was decided that it would be interesting to add other languages to TIIARA as a logical follow-up. The next section describes the translation processes that occurred with eight languages: Arabic, Spanish, Brazilian Portuguese, Mandarin Chinese, Italian, German, Hindi and Russian.

4.0 Translation processes

4.1 Arabic translation

Arabic is not only the mother tongue of millions of people and the most widely spoken language in the Arab world, it is also considered one of the pillars of the Arab nation (Al-Sayyid 1973) and an important indicator (Nishio 2001) of the Arab identity. According to Ethnologue (2015), Arabic is "the largest member of the Semitic branch of the Afro-Asiatic language family" that comprises all descendants of Classical Arabic spoken primarily across the Middle East and North Africa. The Arabic language, with an estimated 223 million speakers throughout the world, is the official or co-official language of 25 countries that include, among others, Algeria, Bahrain, Djibouti, Egypt, Iraq, Jordan, Kuwait, Lebanon, Libya, Mauritania, Morocco, Oman, Palestinian West Bank and Gaza, Qatar, Somalia, Sudan, Syria, Tunisia, United Arab Emirates and Yemen. In addition to the Arab countries, in which Arabic speakers are concentrated, large numbers of Arabic speakers can be found all over the world. However, the Arabic language situation has changed over time because of the influence of foreign and indigenous local languages. For example, in countries that were under British rule, such as Egypt, Jordan and Iraq, English remains strong even today, whereas French remains strong in countries affected by French colonization, such as Algeria, Tunisia, Morocco and Lebanon.

Arabic language can be classified into three main variants: Classical Arabic, which is used in the Holy Quran; Modern Standard Arabic (MSA), which is used in the media, education and modern literature; and Colloquial Arabic, which differs from one country to another and one region to another within the Arab world. Those dialects differ from MSA and each other in terms of phonology, morphology, lexical choice and syntax. The translation of the TIIARA was primarily based on MSA; for example, translating "window" into شباك rather than دريشة, which is used locally in the Gulf.

The Arabic alphabet contains 28 letters. These letters can have four types of diacritical marks that typically represent short vowels and indicate how a word should be pronounced. According to Habash (2010), letters are always written, but diacritics are optional: written Arabic can be fully diacritized, partially diacritized or entirely undiacritized. If these diacritical marks are improperly used or not used at all, this may cause errors in the pronunciation and hence change the meanings of words. Even more challenging, these marks are not usually indexed in online information retrieval systems. An example from TIIARA would be حب ("love"), as it can be written as intended حُب, or as حَب ("grains") when changing the diacritical mark on the first letter.

Another example would be "beds." The translation is أسِرّة, but the diacritics ◌ِ and ◌َّ are needed to differentiate it from أُسرة ("family"), and those marks are not represented in online systems, as stated before. Another example would be "writers." The translation is كُتّاب, but the word could be easily confused with كِتاب ("books") without the proper diacritical marks.

Azmi and Aljafari (2015) noted that some terms related to innovations and borrowed words might not be regulated. Since TIIARA was originally developed simultaneously in English and French, there were a number of terms that are not found in the Arabic language. To overcome this challenge, the terms were transliterated into Arabic. For example, "curling" was transliterated into الكيرلنج, accompanied by the word "sport" رياضة to indicate it is a sport. The same process was used for هوكي الجليد ("hockey"), رياضة الكروس ("lacrosse") and حياكة وكروشيه ("knitting" and "crocheting"). Other foreign terms were directly translated into Arabic, such as مشغلات أقراص فيديو رقمية ("DVD players"), ماسحات ضوئية ("scanners"), مسجل فيديو شخصي ("personal video recorders") and مراكز تجارية ("malls"). However, people tend to use the English term whenever they refer to these inventions.

Habash (2010) notes that MSA orthography has largely been standardized for a long time now. However, few variations persist across and within different Arab countries. In TIIARA, for example, "television" was transliterated into تلفزيون, but others might write the short vowel in the word and it would be تيلفيزيون. "Golf," on the other hand, was transliterated into غولف, but it could also be written as جولف because the /g/ sound does not occur in MSA and is replaced with the closest letters to that sound, غ and ج in this case.

Arabic language has around 5 million words that are derived from around 11,300 roots compared to 400,000 keywords in English, which has total of 1.3 million words. As a result, there is a wealth of terminology and a complex morphology (Habash 2010) compared to English and European languages. Examples from TIIARA included the options to translate "elephants" into أفيال or فيلة, "sheep" into غنم or أغنام, "canyons" into وديان or أودية. The latter translation in these cases was chosen to simplify the process. There was also the question whether the definite article "the" ال (*Al*) should be included or not in the translation. Examples include الناس ("people"), إسلام ("Islam"), المنسوجات ("textiles"). These terms were treated based on the context and not one single rule was followed.

Finally, there were a few instances where the politically correct terms needed to be included rather than using the literal translation of the word. For example, "disabled people" was translated into ذوي الاحتياجات الخاصة ("people with special needs") rather than معاقون, which literally means "handicapped." Another example would be translating "gay people" into مثليون rather than شواذ, which literally means "irregulars."

4.2 Spanish and Portuguese translations

Like the French language, Spanish belongs to the Romance branch of the Indo-European language family. Spanish was developed from Vulgar Latin in an area of the Iberian Peninsula that is now Spain. It was largely exported to the Americas and other parts of the world (the Philippines, parts of Oceania) during the Spanish colonization (Ethnologue 2015) that took place in the 16th century. The Spanish language is used worldwide (Ethnologue 2015) by more than 328 million people as a first language and by some 60 million people as a second language. Spanish is the official language (exclusively or with other languages) of Argentina, Bolivia, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Equatorial Guinea, Guatemala, Honduras, Mexico, Nicaragua, Panama, Peru, Puerto Rico, Spain, Uruguay and Venezuela.

The Spanish translation and revision was realized at a pace of approximately 100 words per hour (seven hours total). For the Spanish version, the English version of TIIARA was used first, and then the French version of the categories was consulted. For example, the term "stove" in English was more in line with "estufa" in Spanish than "four" in French. Multiword expressions have posed difficulties because sometimes a single word in Spanish corresponds to a multiword expression in English or French and vice versa. For example, the English word "garbage" was translated as "sacar la basura," which is closer to the French expression "sortie des ordures." Other words were directly spanishized. For example, "Hanukkah" was converted into "Janucá." In addition, words such as "hockey" or "curling" were not translated, because they are words borrowed from other languages, mostly from English. *Wordreference* dictionary as well as *Wikipedia* were regularly consulted, especially for the translation of words that are not frequently used in Spanish. Furthermore, English verbs ending with the suffix "ing" (e.g., cooking, sleeping, etc.) have been translated as action verbs, unlike the French, which were transformed into substantives.

Similarly to Spanish, Portuguese is another language that belongs to the broad Romance branch of the Indo-European language family. Ethnologue (2015) calculates there is a population of 180 million people worldwide that has Portuguese as a first language. As a result of the important colonial expansion that took place in the 15th and 16th centuries, Portuguese is spoken in Portugal as well as in other countries, such as Brazil, and in the Portuguese colonial and formerly colonial territories of Angola, Cape Verde, East Timor, Goa, Guinea-Bissau, Macau, Mozambique and São Tome e Principe.

Portugal counts 10 million speakers and Brazil has 164 million. It should be noted that important differences exist between the Portuguese spoken in Portugal and the Portuguese spoken in Brazil. For the Brazilian Portuguese translation, the English version was favoured, as well as online dictionaries. The translation and revision were completed in approximately 10 hours. No particular difficulties were reported during the process.

4.3 Mandarin Chinese translation

With almost 850 million speakers (Ethnologue 2015), Mandarin Chinese is the world's largest language. It is the official language of mainland China and Taiwan, serving as a "solution" to the multiplicity of dialects that exist and are used in that portion of the globe. Chinese Mandarin belongs to the Sino-Tibetan language family, which is one of the largest in the world, second to the Indo-European language family in terms of number of speakers. It is used for formal and informal communication. Mandarin Chinese speakers can mainly be found in China and Taiwan, but also in Brunei, Cambodia, Hong Kong, Indonesia, Laos, Malaysia, Mauritius, Mongolia, Philippines, Russia, Singapore, Taiwan, Thailand, and Vietnam. The online dictionaries iciba.com and frdic.com were widely used in the Chinese translations. TIIARA translation in Chinese took about two periods of three hours to translate all the words. Generally, the translation was not difficult to achieve. However, for a native Chinese speaker with basic knowledge of English and French, sometimes the English words were more precise than their corresponding French ones, for instance "Military honours and promotions" and "Événement militaire," where the Chinese translation is based on the English version. Sometimes, there is diversity in the English version, for instance, “writing,” which may mean the “writing skill” (写作) or the “art of writing” (书法), while its French version (“Art de l’écriture”) corresponds perfectly to the Chinese version (书法).

4.4 Italian translation

Similar to Spanish and Portuguese, the Italian language is part of the Romance branch of the Indo-European language family and is a direct descendant of Latin spoken by the Romans and imposed by them on the peoples under their law. It shares many characteristics with other Romance languages. The Italian language (Thomson 2014) is spoken by 57.7 million people in Italy with a total worldwide of 64 million speakers in some 29 countries, mostly Italy but also in some countries such as San Marino, Switzerland, Vatican City and some parts of Croatia and Slovenia. The Italian translation of TIIARA was produced quite easily. It took more or less two periods of one hour each. The words were simple enough to be translated. In general, the English TIIARA was used first, then the French version for confirmation. Some words were more complex to translate. For example, “graduation” is difficult to translate, because this refers to a concept that does not exist in the Italian culture. In addition, words such as “hockey” or “lacrosse” do not translate, because in Italian these words are borrowed from the English language. Dictionaries were not consulted often, except for some words such as “ladders,” “rakes” and so on that are not really used in Italian. Regarding multiword expressions, if an English expression such as “equipment,” “tools” and “supplies” was encountered, often only one word, such as “attrezzature” or “equipaggiamento,” was sufficient in Italian.

4.5 German translation

The German language belongs to the Germanic branch of the Indo-European language family and is considered to be one of the world’s major languages, spoken by approximately 95 million native and 28 million second-language speakers in 40 countries around the world. The official language of Germany (Thomson 2014) (with Danish, Frisian and Sorbian as minority languages) is also used in Belgium, Denmark, France, Hungary, Italy, Luxembourg, Namibia, Romania, Slovakia and Switzerland. German is written using the Latin alphabet with the addition of three vowels with an umlaut: Ää, Öö, Üü and a special letter ß that represents “ss.” The German translation (including searching and editing) took about 15 hours. The starting point of the translation was the English version. More specifically, when a decision needed to be made between the use of plural or singular, the German translation was aligned with the English version, even in the case that the French/Spanish/Italian/Portuguese translations were using the singular forms. The French translation was used for confirmation. Furthermore, the Spanish version was considered in the limit of the translator’s knowledge of that language. When obtaining different translations for the English or French terms, the German translator chose the most appropriate one. The translation process also required the consultation of some tools, including commercial dictionaries, *Wikipedia*, non-commercial web-based dictionaries or translation websites. On some occasions, some German-language websites of stores (e.g., German or doing business in Germany) that offer the same product ranges as the ones that the translator guessed to be the source of the taxonomy were looked up.

Also, it should be noted that Germany is one of the countries where anglicisms are widely accepted and quickly integrated into the common language, especially with regard to technology-related terms. For these terms, the translation was identical (e.g., “laptop”), sometimes orthographically adapted (e.g., “DVD player,” “MP3 player”). In some cases, German words existed and were used to a certain degree. As a consequence, some words received two translations. Both versions were given (e.g., “Drucker oder Printer”), using “oder” (or) to mark the alternative and listing first whichever term was estimated to be the more frequently used. The German word for “mobile phone” is an outlier (“handy”), which is a false anglicism created and used only in Germany. In the end, the translator reported that the translation process would have been much easier if he had a better knowledge of the classification sources of the different domains compiled into the taxonomy. For example, the website of a building supply store, a medical supply company, a tourist sight and a city information system from which the taxonomy terms had been extracted were examined. Without this crucial information, the translator had to make his own assumptions and found equivalent sites in the German language. Although these websites were helpful sources, their retrieval was considered to be a very time-consuming effort.

4.6 Hindi translation

Hindi is part of the Indo-Aryan branch of the Indo-European language family. One of the official languages of India, it is the main language used, among others, in the northern states of Bihar, Chhattisgarh, Delhi, Haryana, Himachal Pradesh, Jharkhand Rajasthan Madhya Pradesh, Uttarakhand and Uttar Pradesh. This language is written with the Devanāgarī alphabet, a descendant of the Brāhmī script used for writing Marathi, Nepali and Sanskrit. It is also closely related to Urdu (Thompson 2014), the main language of Pakistan, which is written with the Arabic script. Used by more than 540 million individuals, the Hindi vocabulary derives from Sanskrit (Everaert 2010), one of the most studied languages in the world. Overall it took approximately 18 hours to complete the translation of TIIARA from English to Hindi. The main reason explaining that amount of time is the lack of knowledge of the translator, even though considered bilingual with Hindi as the mother tongue. This translator usually does not translate words from Hindi to English or English to Hindi on a daily basis and this was a first and completely new experience. The translator used HinKhoj Dictionary App and Google Translate as translation tools. The translator reported encountering difficulties with several words. For example, words for some animals (e.g., "raccoons") were difficult to translate in Hindi because these animals or birds are not found in India and so there are no words for them in Hindi. For the sake of translation, such words were written in Hindi letters so the words sound the same as in English (e.g., "raccoons" as रकूनस). In addition, there were some words related to religion, such as "First Communion," "Hanukkah" and "Thanksgiving," that were difficult to translate. Thus, in general, if a possible word candidate was not found in Hindi, even with the help of an online dictionary and Google Translate, the translator decided to write the words in Hindi letters with sounds similar to English.

4.7 Russian translation

Russian is a Slavic (Slavonic) language of the Indo-European family, one of three languages that constitutes the Eastern branch of the Slavic language group. It is used (Ethnologue 2015) by 137 million people in Russia and by more than 166 million world-wide as their first language. It is one of the six official languages of the United Nations and the most geographically widespread language and the largest native language in Eurasia. It is also the most widely spoken of the Slavic languages. Russian is the official language and the principal cultural language in the Russian Federation. It is the mother tongue and primarily language of communication for the overwhelming majority of people in Russia and one of the official languages in Belarus, Kazakhstan, and Kyrgyzstan. It is also used as a second language in other former republics of the Soviet Union and is a widely spoken language in Ukraine, Moldova, Latvia, and Estonia. Russian is written using a Cyrillic alphabet developed in the 9th-10th centuries. The modern Russian alphabet consists of 33 letters.

The source document for the translation was the English version of the taxonomy, while the French version was used as a confirmation tool to clarify some uncertain terminology. For example, "air lifts" can have multiple meanings when translated into Russian, so the French translation was useful in determining which translation variant is to be chosen. Similar to the German translation and as opposed to the French, Spanish, Italian, and Portuguese translations, the choice of the singular vs. plural forms was aligned with the English original. Similar to German, anglicisms are readily adopted in Russian (usually through phonetic borrowing) for the terms denoting new technological developments. The examples from the TIIARA taxonomies are "Emoticons/ Эмотиконы" or "Gyropods/ Гироподы" that have been phonetically borrowed from English but grammatically adapted in Russian, i.e. received Russian affixes (endings). Other examples of phonetically borrowed but in this case not grammatically adapted terms include the realities and customs non-endemic to Russian culture, such as names of religious holidays, e.g. "Diwali /Дивали."

The Russian translation was performed combining the methods of machine translation with the traditional human-made translation. The first draft of the Russian version of the TIIARA was created using the machine translation software PROMT 7 Professional that uses the hybrid technology that combines statistical and rule-based translation methodologies. The machine-created first draft of the translation was subsequently edited and corrected by a human translator. The process that involved extensive term-searching was done using the commercial electronic desktop-based dictionary, *Lingvo x6* that includes 69 Russian-English-Russian general and specialised dictionaries with the total of over 3,200,000 entries, *Wikipedia*, and *Wiktionary*. In total, the process of translation, including creation of the first machine-translated draft, term searching, subsequent editing and proofreading took close to 10 hours.

5.0 Discussion

According to Lambe (2007), a taxonomy is a "structured set of names and descriptions used to organize information and documents in a consistent way." A taxonomy is considered to be a loosely controlled vocabulary, where relations do not necessarily follow traditional patterns encountered in a thesaurus (equivalence relations, hierarchical relations and associative relations). In the case of TIIARA, the number of top-level categories as well as the depth of the taxonomic structure was restricted to seven main categories. This number is close to the nine top-levels facets reported regularly (Lambe 2007) in the literature. It is also generally recommended to restrict the level of depth to four, developing as many as five in some instances. Also, it is worth mentioning that TIIARA started as a bilingual vocabulary (English and French), that is, categories and subcategories were simultaneously developed in both languages in order to keep the structure as parallel as possible. Nevertheless, several linguistic questions arose in the development of the bilingual vocabulary. For example, the development of the taxonomy took place in Quebec, one of the provinces of Canada. In Canada, most French speakers live in Quebec. The French language, one of the two official languages of Canada, is used by the majority of Quebecers. The French language spoken in Quebec is not all that different from the language used in France, but some variations exist because of regional languages and folk dialects spoken in France at the time of colonization. However, it is worth mentioning that the observed differences are not important at the terminological level, with some exceptions. Most divergences occurred rather at the syntax level, since French-Canadians are constantly exposed to and have a tendency to mimic English-language structures. Consequently, even if the choice of French terms used for TIIARA categories and subcategories was not that complex, great care was taken to keep the French terms as "international" as possible. It is also worth mentioning that the English counterpart of TIIARA was defined by two Americans. This also opened the door to several terminological choices oriented toward an "international" English rather than Canadian English, American English or British English.

Countries whose populations have diverse languages face significant problems with multilingualism. Notably, this is the case for Belgium, Canada and Switzerland. In a larger context, it is also the case for the European Union. Other countries, such as China, for example, somehow mitigated this problem since they adopted a single writing system. Nevertheless, access to multilingual information has become more and more important over the last decade. It has also implied a reorganization of Internet services, including the indexing languages and the functionalities of search engines. This observation led to the translation of TIIARA in multiple languages. A quick survey of the existing controlled vocabularies revealed that not many of them could claim to be multilingual, let alone bilingual. For example, the *Library of Congress Subject Headings* (*LCSH*) is a quasi-thesaurus from which the subject of library documents (books, articles, websites, etc.) is selected. It is an accumulation of the headings established at the U.S. Library of Congress since 1898. It currently contains (Library of Congress 2011) more than 220,000 English terms. The Multilingual Access to Subjects (MACS) project aims to provide multilingual access to subjects in the catalogues of the participants. This project was conceived by the Deutsche Bibliothek (Schlag-WortnormDatei), the British Library (*Library of Congress Subject Headings*), the Bibliothèque nationale de France (Répertoire d'autorité-matière encyclopédique et alphabétique unifié) and the Swiss National Library. It is worth mentioning that the RAMEAU language has been developed since 1980 in an autonomous way at Quebec's Université Laval. The "Répertoire de vedettes-matière" (Université Laval 2011) is a translation/adaptation of the *Library of Congress Subject Headings*. Some English and French equivalents already exist (Landry 2006), and this allows a search of some French library catalogues with the *LCSH*. In the field of iconographic description, Iconclass is a specific international classification that museums can employ for iconographic research and the documentation of images. It contains definitions (approximately 28,000) of objects, people, events, situations and abstract ideas that can be the subject of an image. The Iconclass browser offers English, German, French and Italian keywords, as well as descriptions. Partial translations (Iconclass 2011) in Finnish and Norwegian and experimental translations in Chinese and Dutch also exist. Nevertheless, real multilingual controlled vocabularies are very scarce. According to the Working Group on Guidelines for Multilingual Thesauri (IFLA 2009):

> There are three approaches in the development of multilingual thesauri: (1) Building a new thesaurus from the bottom up: (a) starting with one language and adding another language or languages; (b) starting with more than one language simultaneously; (2) Combining existing thesauri: (a) merging two or more existing thesauri into one new (multilingual) thesaurus to be used in indexing and retrieval; (b) linking existing thesauri and subject heading lists to each other; using the existing thesauri and/or subject heading lists both in indexing and retrieval; (3) Translating a thesaurus into one or more other languages.

This third option has been selected for the transition of TIIARA toward multilingualism. The taxonomy now exists in multiple languages (French, English, Arabic, Spanish, Brazilian Portuguese, Italian, Mandarin Chinese, German, Hindi and Russian). Other translations are also being considered (Polish, Japanese, Greek, etc.).

All translations were processed on a volunteer basis, by several translators with several degrees of expertise. In other words, the individuals that participated in the TIIARA translation process cannot be labelled as "professional translators." The main criterion for the selection of a translator was his or her knowledge of the source languages (English or French) and the target languages (one of the aforementioned languages). Two general rules of thumb were taken into consideration (Howe 2008): 1) keeping in mind most participants have just a limited amount of time available; and 2) understanding why people want to get involved and try to match their personal motivations with what you ask them to do. The translation processes described earlier constitutes one of the many examples of the crowdsourcing phenomenon that now exists. According to Howe (2008), "'Crowdsourcing' is the act of taking a task traditionally performed by a designated agent (such as an employee or a contractor) and outsourcing it by making an open call to an undefined but large group of people." In other words, crowdsourcing allocates different people to contribute to a common goal by accomplishing similar tasks, in a cheaper fashion. The benefits of crowdsourcing are easily understandable. It is highly scalable, the workforce is available instantly and as many or as few resources can be obtained as required. Crowdsourcing has been applied to several types of activities (crowdvoting, crowdindexing, crowdfunding) and many domains (history, art history, journalism, health care, creative works, astronomy, genealogy, ornithology, etc.). Several examples of success stories have appeared over the last decade. Among these, it is worth mentioning, *Amazon Mechanical Turk* (MTurk) (Amazon 2015), a crowdsourcing Internet marketplace that allows individuals to coordinate the use of human intelligence to perform tasks that computers are currently unable to do; *Wikipedia* (Wikipedia 2015), an online encyclopaedia that receives entries (created, edited and fine-tuned) from users; *YouTube* (YouTube 2015), a vast collection of video clips submitted by users; *Threadless.com* (Threadless 2015), which allows people to submit T-shirt design ideas that others can then vote on; *Google* (Google 2015), a search engine that integrates all user-generated factors into its web page ranking algorithm; *iStock* (*iStockphoto* now owned by *Getty Images*) (Getty Research Institute 2015b), a web resource for crowdsourced, royalty-free stock images, media and design elements, and which offers millions of hand-picked photos, illustrations, videos and audio tracks.

Crowdsourcing is now well beyond the testing stage. It is currently a workable method of expanding global reach and making content available to individuals in their languages all over the world. In the case of TIIARA, crowdsourcing—or more precisely, crowdtranslation—presents an interesting solution for transforming a bilingual tool into a multilingual tool. Even if automatic translation systems have evolved over the years to become almost reliable tools, relatively speaking, the decision to use human translators was driven by the fact that TIIARA contains only a few words, more or less 2000, included in all four levels. Additionally, because the taxonomic structure is mostly single or multiple words and not complete sentences, the translation by non-expert translators should have been an easy process. However, sometimes words alone do not provide sufficient context to be well interpreted (Kishida 2005; Gey et al. 2005) and, therefore, could be incorrectly translated. Indeed, translators often face many difficulties (Braschler 2004), mainly semantic and syntactic ambiguities. In the context of TIIARA translations, the semantic ambiguity, mainly due to a lack of context, may result in a misinterpretation and, consequently, in an incorrect translation.

It is worth mentioning that the translators could not be considered as "professional taxonomists" either. In other words, they had little knowledge of a controlled vocabulary structure such as a taxonomy while the development of TIIARA in English and French was conducted by expert taxonomists. This lack of expertise could have affected some decisions in the choice of words. According to the Working Group on Guidelines for Multilingual Thesauri (IFLA 2009), "Inter-language equivalence has three aspects: semantic, cultural and structural. The semantic and cultural aspects refer to the meaning of the terms and the way the terms are used in a given language or culture. The structural aspect refers to the hierarchical and associative relations among terms." Thus, if translators could guarantee that the semantic and cultural aspects of the translation were preserved, the situation would be different with regard to the structural dimension. During the TIIARA translation into multiple languages, several scenarios were observed and reported by the translators: 1) some terms were exactly translated (inter-language synonymy); 2) some translated terms were inexact or near equivalent (inter-language quasi-synonymy, with a difference in viewpoint); 3) some translated terms were partially equivalent (inter-language quasi-synonymy, with a difference in specificity); 4) some translated terms corresponded to one-to-many equivalents (too many or not enough terms), where to express the meaning of the preferred term in one of the languages, two or more preferred terms were needed in the other language; and, 5) some terms from the source language did not have equivalent terms in the target language. In the case of TIIARA, this preceding situation was not really acceptable because the taxonomy aimed to be as symmetrical as possible. Consequently, in the case of the latter scenario, the translators felt that the only solution was to borrow the term from the source language to become the equivalent in the target language. This tradition of word exchange between languages has a long history and could be considered (or not) as a good practice by the so-called protectors of the "lingua franca." However, several examples of borrowed terms have been integrated into languages for years and this "better than nothing" solution is not considered to be heretic anymore.

6.0 Conclusion

TIIARA's data is contained in a single file, complying with the Extensible Markup Language (XML) web standard. Characters comply with the Universal Character Set Transformation Format (UTF-8). This ensures that every image category can be described in potentially all existing languages. In addition, the structure of TIIARA's XML file corresponds to a single table in a relational database: Category (id : integer, belongsto : integer, level : integer, langage_1 : string, ... , langage_n : string). Each image category is identified by a unique number (id) and linked to a wider category (belongsto). It is thus also possible to distinguish two categories that share a unique name by listing their parent categories. For instance, there is one category named "Death," whose path is "Daily Life"> "Event">"Negative Event">"Death" and a second one whose path is "Abstract Ideas">"Concepts">"Death." The field named level is an integer that gives the length of such paths. For the above examples, Death is at level 4 as an Event and level 3 as a Concept. This simple and conventional data structure and format allow the easy import and export of TIIARA's data into any SQL database and therefore take advantage of a powerful querying language strongly connected to web technologies. Table 2 presents an example of one TIIARA term in 10 languages coded in XML.

```
<row>
<column name="id">475</column>
<column name="english">Squirrels</column>
<column name="belongsto">193</column>
<column name="french">Écureuil</column>
<column name="spanish">Ardilla</column>
<column name="chinese">松鼠</column>
<column name="portuguese">Esquilos</column>
<column name="arabic">سناجب</column>
<column
name="german">Eichhörnchen</column>
<column name="hindi">गिलहरीयाँ</column>
<column name="italian">Scoiattolo</column>
<column name="russian">белки</column>
<column name="level">4</column>
</row>
```

Table 2. Excerpt from TIIARA

Multilingual information processing has gained more and more attention in recent years. The TIIARA taxonomy has been tested and validated (Ménard and Dorey 2014) by a number of participants during several evaluation processes. At each stage of the testing process, the taxonomy was refined until there was a satisfactory level of consistency. It was of great importance that the taxonomy be able to evolve as users discovered its restrictions. However, in order to create stability, it was decided that TIIARA will not be amended during the translation processes. Now completed, updates in multiple languages will be added regularly. The bilingual version formed the basis of an intuitive and accessible database. The next step of this research will be the testing of the translated versions with real image searchers. Another project will be to include the multilingual TIIARA in SINCERITY (Search INterfaCE for the Retrieval of Images with a TaxonomY), a bilingual search engine that has been developed in parallel with the present project in order to offer members of other linguistic communities equivalent opportunities and subsequently bridge the information divide that still exists. SINCERITY allows users to search for images (from the database IDOL constructed for the project) with a traditional search box, where a query is formulated using the users' own words. However, SINCERITY (Ménard et al. 2013) also offers the possibility to initiate the queries and search the database using the TIIARA structure (in English or French). Even if very few search engines provide a taxonomic structure to initiate queries, browsing a subject classification system could be a useful functionality, especially for image searchers who have difficulties formulating a query with their own words. Adding more languages in TIIARA and SINCERITY will constitute an added value for the image searchers, who are often limited by their lack of knowledge of multiple languages.

Notes

1. The complete presentation of the multilingual vocabularies examined in the preliminary phase of this research can be found in Ménard and Smithglass 2012.
2. The complete presentation of the TIIARA taxonomy development can be found in Ménard 2012 and 2013.

References

ABBYY. 2013. *Abbyy Lingvo-Online Dictionary*. http://www.lingvo-online.ru/en.

Aitchison, Jean, David Bawden and Alan Gilchrist. 2000. *Thesaurus Construction and Use: A Practical Manual*. London: Aslib IMI.

Al-Sayyid, Jalal. 1973. *Haqiqat al-Umma al-Arabiyya wa awamil hifziha wa tamziqiha*. Beirut: Dar al-Yaqza al Arabiyya.

Azmi, Aqil M. and Eman A. Aljafari. 2015. "Modern Information Retrieval in Arabic - Catering to Standard and Colloquial Arabic Users." *Journal of Information Science* 41: 506-17.

Amazon. 2015. *Amazon Mechanical Turk*. https://www.mturk.com/mturk/welcome.

Braschler, Martin. 2004. "Combination Approaches for Multilingual Text Retrieval." *Information Retrieval* 7: 183-204.

Everaert, Christine. 2010. *Tracing the Boundaries Between Hindi and Urdu: Lost and Added in Translation between 20th Century Short Stories*. Leiden; Boston: Brill.frdic.com. 2015. *Online Dictionary.* http://frdic.com/.

Getty Research Institute. 2015a. *Art & Architecture Thesaurus® Online*. http://www.getty.edu/research/tools/vocabularies/aat/

Getty Research Institute. 2015b. *iStock by Getty Images*. http://www.istockphoto.com

Gey, Frederic C., Noriko Kando and Carol Peters. 2005. "Cross-Language Information Retrieval: theWay Ahead." *Information Processing & Management* 41: 415-31.

Google. 2015. *Google*. http://www.google.com.

Habash, Nizar Y. 2010. *Introduction to Arabic Natural Language Processing*. San Rafael, Calif.: Morgan & Claypool Publishers.

Howe, Jeff. 2008. *Crowdsourcing – Why the Power of the Crowd is Driving the Future of Business*. http://www.bizbriefings.com/Samples/IntInst%20---%20Crowdsourcing.PDF.

Iciba. 2015. *Iciba Online Dictionary*. http://www.iciba.com/

Iconclass. 2015. *What is Iconclass?* http://www.iconclass.nl/home

International Federation of Library Associations and Institutions – IFLA. 2009. *Guidelines for Multilingual Thesauri*. http://www.ifla.org/files/assets/hq/publications/professional-report/115.pdf

Jiménez-Crespo, Miguel A. 2013. *Translation and Web Localization*. New York: Routledge.

Jörgensen, Corinne. 2003. *Image Retrieval - Theory and Research*. Lanham: Scarecrow Press.

Kishida, Kazuaki. 2005. "Technical Issues of Cross-Language Information Retrieval: A Review." *Information Processing & Management* 41: 433-55.

Lambe, Patrick. 2007. *Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness*. Oxford: Chandos Publishing.

Landry, Patrice. 2006. "Multilinguisme et Langages Documentaires: Le Projet MACS en Contexte Européen." *Documentation et Bibliothèques* 52: 121-9.

Library of Congress. 2015a. *Library of Congress Subject Heading List*. http://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html#About

Library of Congress. 2015b. *Thesaurus for Graphic Materials*. http://www.loc.gov/pictures/collection/tgm/

Ménard, Elaine. 2013. "TIIARA for an IDOL: Indexing Adventure of a Small Digital Image Collection." *The Indexer* 31, no. 1: 2-11.

Ménard, Elaine. 2012. "TIIARA: The 'Making of' a Bilingual Taxonomy for Retrieval of Ordinary Images." *Library Hi Tech* 30, no. 4: 643-54.

Ménard, Elaine and Jonathan Dorey. 2014. "TIIARA: A New Bilingual Taxonomy for Image Indexing." *Knowledge Organization* 41: 113-22.

Ménard, Elaine and Vanessa Girouard. 2015. "Image Retrieval with SINCERITY: A Search Engine Designed for Our Multilingual World!" *OCLC Systems and Services* 31, no. 4: 204-18.

Ménard, Elaine and Margaret Smithglass. 2014. "Digital Image Access: An Exploration of The Best Practices of Online Resources." *Library Hi Tech* 32, no. 1: 98-119.

Ménard, Elaine and Margaret Smithglass. 2012. "Digital Image Description: A Review of Best Practices in Cultural Institutions." *Library Hi Tech* 30, no. 1: 291-309.

Ménard, Elaine, Nouf Khashman and Jonathan Dorey. 2013. "Two Solitudes Revisited: A Cross-Cultural Exploration of Online Image Searchers Behaviours." In *Human-Centred Design Approaches, Methods, Tools, and Environments: Proceeding of 15th International Conference on Human-Computer Interaction, 21-26 July 2013, Las Vegas, NV*, edited by Masaaki Kurosu. Berlin: Springer, 79-88.

Nishio, Tetsuo. 2001. "Language Nationalism and Consciousness in the Arab World." In *Cultural Change in the Arab World,* edited by T. Nishio, 137-46. Osaka: National Museum of Ethnology.

Thompson, Irene. 2014. *About World Languages!* http://aboutworldlanguages.com/

Threadless. 2015. *Threadless*. https://www.threadless.com/

UNESCO. 2015. *UNESCO Thesaurus*. http://databases.unesco.org/thesaurus/.

Université Laval. 2011. *Répertoire de vedettes-matière*. https://rvmweb.bibl.ulaval.ca/.

YouTube. 2015. *YouTube*. https://www.youtube.com/.

Walter, Henriette. 2001. *Honni soit qui mal y pense*. Paris: Robert Laffont.

Wiktionary. 2015. *Wiktionary, the Free Dictionary*. https://en.wiktionary.org.

Wikipedia. 2015. *Wikipedia the free Encyclopaedia*. https://fr.wikipedia.org.